Statistical Papers 29, 191–203 (1988)

Statistical Papers
Statistische Hefte

An operative extension of the likelihood ratio test from fuzzy data

M.A. Gil and M.R. Casals

Received: August 27, 1987; revised version: March 8, 1988

In the present paper we are going to extend the likelihood ratio test to the case in which the available experimental information involves fuzzy imprecision (more precisely, the observable events associated with the random experiment concerning the test may be characterized as fuzzy subsets of the sample space, as intended by Zadeh, 1965). In addition, we will approximate the immediate intractable extension, which is based on Zadeh's probabilistic definition, by using the minimum inaccuracy principle of estimation from fuzzy data, that has been introduced in previous papers as an operative extension of the maximum likelihood method.

Keywords: fuzzy information, likelihood ratio test, maximum likelihood estimation, minimum inaccuracy estimation, Zadeh's probabilistic definition.

1. INTRODUCTION

Experiments are going to be considered in which the person responsible for observation cannot always crisply perceive their outcomes, but each *observable elementary event* may only be assimilated with a *fuzzy subset* of the sample space (Zadeh, 1965) or, more precisely, with fuzzy information, as intended by Tanaka, Okuda and Asai (1978, 1979), and Zadeh (1978), that is defined as follows:

Let $\mathbf{X}$ be a *random experiment* characterized by a probability space (X, β_X, P_θ), where P_θ belongs to a specified parametric family of probability measures $\{P_\theta, \theta \in \Theta\}$ on (X, β_X). We hereafter assume that the sample space X is a set in a euclidean space (usually $\mathbb{R}$) and β_X is the smallest Borel σ-field on X. In addition, the parameter space Θ is a set in a euclidean space so that the unknown parameter θ is numerical or vector-valued.

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

The scheme in Figure 1 explains the mechanism that leads to the obtention of fuzzy information according to the notation in Definition 1.1.

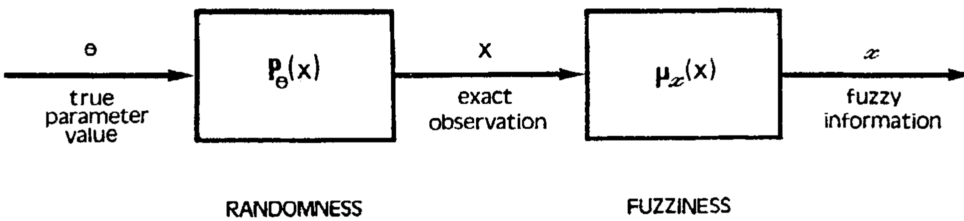

Fig. 1. Process leading to fuzzy information associated with a random experiment

The grade of membership $\mu_{\mathcal{x}}(x)$ describing the fuzzy information $\mathcal{x}$ is often interpreted as a kind of "probability with which the person responsible for observation perceives $\mathcal{x}$ when x is the true experimental outcome". On the basis of this interpretation Tanaka *et al.* (1979) define the *collection of all available observable elementary events* as an orthogonal system associated with **X**, that is

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

$$\sum_{\mathcal{x} \in \mathcal{X}} \mu_{\mathcal{x}}(x) = 1 \qquad \text{for all} \quad x \in X$$

It should be emphasized that $\mathcal{X}$ is an ordinary set whose elements are fuzzy subsets. For this reason, we could easily construct a σ-field on $\mathcal{X}$ (e.g., parts of $\mathcal{X}$).

In order to state a probabilistic model for the random experiments involving fuzzy imprecision we can now introduce the probability of a fuzzy event (Zadeh, 1967).

Definition 1.3. The *probability distribution on* $\mathcal{X}$ *induced by* P_θ is the mapping $\mathcal{P}_\theta$ from $\mathcal{X}$ to [0,1] given by

$$\mathcal{P}_\theta(\mathcal{x}) = \int_X \mu_{\mathcal{x}}(x)\, dP_\theta(x) \qquad \text{for all} \quad \mathcal{x} \in \mathcal{X}$$

Remark 1.1: When we adopt the procedures we are now going to develop, the orthogonality condition assumed for the set of all available fuzzy observations from the experiment is not a strong constraint (in other words, the orthogonality condition does not mean loss of generality, and it will considerably simplify the extension in Section 2). This circumstance is confirmed in greater detail later (Remark 3.3).

Consequently, although the probabilistic framework is not enough by itself to provide us with a suitable model, the Theory of Fuzzy Sets complements the Probability Theory and allows

On the basis of these arguments and concepts we have previously extended methods to solve some statistical problems with imprecise data (1984a, 1984b, 1985a, 1985b, 1986a, 1986b, 1987a, 1988a, 1988b). In particular, we have developed (1986a) the extension of the Neyman-Pearson method of testing simple statistical hypotheses and the Bayes method for composite hypotheses (1986b) from fuzzy data.

In the same way, we are now going to extend the likelihood ratio test concerning composite statistical hypotheses, when the available experimental information is fuzzy.

If a simple random sampling of size n from the experiment $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, is considered and the ability to observe does not permit one to perceive exactly the experimental outcomes, the following notions (1984a, 1984b, 1985a, 1985b, 1986a, 1986b, 1987a, 1988a, 1988b) supply an operative model to express the available sample observations with fuzzy imprecision:

Let $\mathbf{X}^{(n)} = (X^n, \beta_{X^n}, P_\theta)$, $\theta \in \Theta$, be a simple random sample of size n from $\mathbf{X}$, and let $\mathcal{X}$ be a f.i.s. associated with $\mathbf{X}$.

Definition 1.4. An n-tuple of elements in $\mathcal{X}$, $(x_1, \ldots, x_n)$, representing the algebraic product of $x_1, \ldots,$ and x_n, is called *sample fuzzy information of size* n *from* $\mathcal{X}$ (where (x, x') = algebraic product of x and x', with $\mu_{(x,x')}(x,x') = \mu_x(x)\,\mu_{x'}(x')$).

Definition 1.5. A *fuzzy random sample of size* n *from* $\mathcal{X}$, $\mathcal{X}^{(n)}$, (*associated with the random sample* $\mathbf{X}^{(n)}$) is the set consisting of all algebraic products of n elements in $\mathcal{X}$

Remark 1.2: It is worth emphasizing 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 $(x_1, \ldots, x_n)$ would be given by the expression $\mu_{(x_1,\ldots,x_n)}(x_1,\ldots,x_n) = f(\mu_{x_1}(x_1), \ldots, \mu_{x_n}(x_n), x_1, \ldots, x_n)$, 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 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 the Zadeh's sense, 1967) of the fuzzy observations from them, whenever f is the product above.

2. AN IMMEDIATE EXTENSION OF THE LIKELIHOOD RATIO TEST WITH FUZZY DATA

Consider the random experiment $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, and a f.i.s. $\mathcal{X}$ associated with $\mathbf{X}$. Let $(x_1, \ldots, x_n)$ be a sample fuzzy information from $\mathcal{X}$, and let $\nu(x)$ denote the observed absolute frequency of the fuzzy information x in the considered sample fuzzy information. Then

Definition 2.1. The expression

$$\mathcal{L}_\theta = \mathcal{L}(x_1, \ldots, x_n; \theta) = \prod_{x \in \mathcal{X}} [\mathcal{P}_\theta(x)]^{\nu(x)}$$

is called the *likelihood function of* θ *for* $(x_1, \ldots, x_n)$.

Definition 2.2. A parameter value $\hat{\theta}(x_1, ..., x_n) \in \Theta$ satisfying the relation

$$\mathcal{L}(x_1, ..., x_n ; \hat{\theta}(x_1, ..., x_n)) = \max_{\theta \in \Theta} \mathcal{L}(x_1, ..., x_n ; \theta)$$

is called *maximum likelihood estimate of* θ *in* Θ *for the sample fuzzy information* $(x_1, ..., x_n)$.

As the adopted modelization permits us to regard mathematically a random experiment providing fuzzy information as a discrete experiment containing r possible outcomes (r = cardinality of $\mathcal{X}$) the immediate *extended likelihood ratio test* is given by

Theorem 2.1. *Let* $\mathcal{X}^{(n)}$ *be a fuzzy random sample of size* n *from the available f.i.s.* $\mathcal{X}$ *on the random experiment* $\mathbf{X}$, *and assume that the parametric probability measure on* $\mathbf{X}$ *determines a parametric distribution function that is regular in all of its first and second* θ*-derivatives* ($\theta = (\theta_1,...,\theta_k)$ *being a k-dimensional parameter*). *The test consisting in rejecting the null hypothesis* $H_0 : \theta \in \Theta_0 = \{ (\theta_1,...,\theta_k) \in \Theta \mid \theta_1 = \theta_1^\circ, ..., \theta_s = \theta_s^\circ \}$, $(s \leq k)$, *against the alternative* $H_1 : \theta \in \Theta - \Theta_0$, *when* $-2 \log \Lambda_n > c$, *where* Λ_n *is the statistic based on* $\mathcal{X}^{(n)}$ *such that*

$$\Lambda_n = \prod_{x \in \mathcal{X}} [\mathcal{P}_{\hat{\theta}_0}(x) / \mathcal{P}_{\hat{\theta}}(x)]^{\nu(x)}$$

($\hat{\theta}_0$ = *maximum likelihood estimate of* θ *in* Θ_0 *for the obtained sample fuzzy information,* $\hat{\theta}$ = *maximum likelihood estimate of* θ *in* Θ *for the obtained sample fuzzy information,* $\nu(x)$ = *observed frequency of* x *in the obtained sample fuzzy information*), *and* c *is the* 100α % *point of the chi-square distribution with* s *degrees of freedom is a test at a significance level close to* α *as the sample size* n *is large* (*more precisely, when* H_0 *is true, the asymptotic distribution of* $-2 \log \Lambda_n$ *is* χ^2_s). ●

Nevertheless the preceding theorem supplies a procedure that becomes often unmanageable in practice because of the difficulties in determining the maximum likelihood estimate of a parameter from a sample fuzzy observation $(x_1, ..., x_n)$ on the basis of Zadeh's probabilistic definition (and, frequently, because of the difficulties in expressing the likelihood function of θ for the sample fuzzy information so that its logarithm is easy to derive with respect to θ). However, the following example encloses a situation in which the maximum likelihood estimation from fuzzy data is easy to apply:

Example 2.1: Certain pieces are manufactured by a two-staged process so that in the first stage 5 % of pieces have a specific flaw E and in the second one 5 % of pieces have another specific flaw F. Assume that the presence of both, E and F, in a piece determines its defectiveness, and let θ denote the unknown fraction defective.

If the mechanisms of detection of E and F are exact, let $\mathbf{X}$ be the random experiment in which a piece having E is inspected for detection of F. Then, the (conditional given E) probabilities associated with $\mathbf{X}$ are

$$P_\theta(1) = 20\,\theta, \quad P_\theta(0) = 1 - 20\,\theta, \quad \theta \in \Theta = [0, 0.05]$$

(where we consider x=1 if the inspected piece has F,and x=0 if the inspected piece has not F).

Suppose now that the mechanism for detecting E is exact but that for detecting F is not exact. In particular, assume that the inspection of each piece for presence of F only allows us to distinguish between the vague observations f = "the piece seems to have F", $\bar{f}$ = "the piece seems not to have F (or to have $\bar{F}$)", that we describe by means of the following membership functions:

$$\mu_f(1) = 0.75, \quad \mu_f(0) = 0.25, \quad \mu_{\bar{f}}(1) = 0.25, \quad \mu_{\bar{f}}(0) = 0.75$$

Then, the fuzziness in the available information from $\mathbf{X}$ leads to the f.i.s. $\mathcal{X} = \{ f, \bar{f} \}$ whose induced distribution is given by

$$\mathcal{P}_\theta(f) = 10\,\theta + 0.25, \quad \mathcal{P}_\theta(\bar{f}) = 0.75 - 10\,\theta \ , \theta \in \Theta = [0, 0.05]$$

In order to test the null hypothesis H_0: $\theta = \theta_0 = 0.025$, n=90 pieces having the flaw E are independently examined for detection of the flaw F and provide us with the observed frequencies $\nu(f) = 36$, $\nu(\bar{f}) = 54$. According to Definition 2.2 , the likelihood estimate is $\hat{\theta}$ = 0.015, so that

$$\mathcal{P}_{\theta_0}(f) = 0.5, \quad \mathcal{P}_{\theta_0}(f) = 0.5, \mathcal{P}_{\hat{\theta}}(\bar{f}) = 0.4, \quad \mathcal{P}_{\hat{\theta}}(\bar{f}) = 0.6$$

whence $-2 \log \Lambda_n = 3.624 < \mathbf{c} = 3.84$ = 5 % point of the chi-square distribution with s=k=1 degree of freedom. Consequently, the extended likelihood ratio test would lead to non rejection of the hypothesis H_0 at the significance level $\alpha = 0.05$.

When to solve the likelihood equation is extremely difficult, it is worth recalling that when we try to apply the maximum likelihood method from grouped experimental observations (that may be regarded as particular examples of fuzzy information) one must appeal to obtain approximate solutions under certain natural assumptions concerning the choice of the grouping (Cramér, 1946). In the same way it would be interesting to try to approximate the maximum likelihood estimation from sample fuzzy information under some plausible requirements.

In the following section we will verify that the use of the inaccuracy measure defined by Kerridge (1961) allows us to extend the maximum likelihood principle of estimation and this extension provides us with a manageable approximation of the trivial extension gathered in Definition 2.2.

3. AN OPERATIVE APPROXIMATION OF THE LIKELIHOOD RATIO TEST WITH FUZZY DATA

As we have remarked the main inconvenience in applying the immediate extension of the likelihood ratio test with fuzzy data in Theorem 2.1 lies in the lack of operativeness in the extension of the maximum likelihood principle of estimation proposed in Definition 2.2.

It should now be pointed out that the operativeness of the maximum likelihood method based on exact experimental data is fundamentally caused by the introduction of logarithms (because of the form of most of the parameter families of distributions). However, the direct use of logarithms in $\mathcal{L}(x_1, \ldots, x_n ; \theta)$ obviously becomes intractable. On the basis of this argument we have suggested in previous papers (1984a, 1987b, 1988b) the procedure we

Definition 3.1. Let $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, be an experiment where $\{P_\theta, \theta \in \Theta\}$ is a parametric family of probability measures which is dominated by the counting or the Lebesgue measure (generically denoted by λ). Assume that the set $\{(x_1,...,x_n) \mid L(x_1,...,x_n;\theta) > 0\}$, which will be denoted by X^n, does not depend on θ, $L_\theta = L(x_1,...,x_n;\theta)$ being the likelihood function of θ for the exact sample point $(x_1,...,x_n)$. A *minimum inaccuracy estimate of* θ *for the sample fuzzy information* $(\mathcal{x}_1, ..., \mathcal{x}_n)$ is a value $\theta^*(\mathcal{x}_1, ..., \mathcal{x}_n) \in \Theta$ satisfying that

$$\mathfrak{I}(\mu_{|(\mathcal{x}_1,...,\mathcal{x}_n)|}; L_{\theta^*(\mathcal{x}_1,...,\mathcal{x}_n)}) = \min_{\theta \in \Theta} \mathfrak{I}(\mu_{|(\mathcal{x}_1,...,\mathcal{x}_n)|}; L_\theta)$$

where

$$\mathfrak{I}_\theta = \mathfrak{I}(\mu_{|(\mathcal{x}_1,...,\mathcal{x}_n)|}; L_\theta) = -\int_{X^n} \mu_{|(\mathcal{x}_1,...,\mathcal{x}_n)|}(x_1,...,x_n) \log L(x_1,...,x_n;\theta)\, d\lambda(x_1) \ldots d\lambda(x_n)$$

(with $|(\mathcal{x}_1, ..., \mathcal{x}_n)|$ the "standardized form" of $(\mathcal{x}_1, ..., \mathcal{x}_n)$ as defined by Saaty, 1974, that is, $\mu_{|\mathcal{x}|}(\cdot) = \mu_{\mathcal{x}}(\cdot)/\underline{\lambda}(\mathcal{x})$, where $\underline{\lambda}(\mathcal{x}) = \int_X \mu_{\mathcal{x}}(x)\, d\lambda(x)$).

Remark 3.1: The minimum inaccuracy principle was exhaustively justified in a previous paper (1984a). It should be emphasized that it is equivalent to minimize the expression $\mathfrak{I}_\theta$ with respect to θ to minimize

$$-\int_{X^n} \mu_{(\mathcal{x}_1,...,\mathcal{x}_n)}(x_1,...,x_n) \log L(x_1,...,x_n;\theta)\, d\lambda(x_1) \ldots d\lambda(x_n),$$

so that, we have only considered the standardized version of $(\mathcal{x}_1, ..., \mathcal{x}_n)$ to obtain the well-known "measure of inaccuracy" defined by Kerridge (1961) between the membership function of $|(\mathcal{x}_1, ..., \mathcal{x}_n)|$ and the likelihood function L_θ (which is only defined for density functions). On the other hand, it is worth pointing out an interesting analogy between the new suggested extension and the preceding one: the immediate extension of the maximum likelihood principle (Definition 2.2) would reduce to minimize the inaccuracy of order β=2 (Rathie and Kannappan, 1973, Mathai and Rathie, 1975) between the membership function of $|(\mathcal{x}_1, ..., \mathcal{x}_n)|$ and the likelihood function L_θ.

The operativeness of the minimum inaccuracy principle is corroborated in the following theorem (Corral *et al.*, 1984a, Gil *et al.*, 1988b) which establishes a constructive way to easily obtain the minimum inaccuracy estimates from the efficient (and, consequently, from the maximum likelihood) ones based on exact information.

Theorem 3.1. *Let* $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, *be an experiment where* $\{P_\theta, \theta \in \Theta\}$ *is a parametric family of probability measures which is dominated by the counting or the Lebesgue measure* λ. *Assume that*

i) Θ *is an interval in a euclidean space.*

ii) $X^n = \{(x_1,...,x_n) \mid L(x_1,...,x_n;\theta) > 0\}$ *does not depend on* θ.

iii) P_θ *determines a parametric distribution function that is regular in all of its first and second* θ*-derivatives.*

iv) $\underline{\lambda}(x_1, \ldots, x_n) = \prod_{x \in \mathcal{X}} [\underline{\lambda}(x)]^{\nu(x)} < \infty$.

If $\mathbf{T}$ *is an estimator of* θ *based on a (nonfuzzy) random sample from* $\mathbf{X}$ *and attaining the Fréchet–Cramér–Rao lower bound for the variance, and with expected value* $\mathbf{h}(\theta)$ *where* $\mathbf{h}$ *is a one-to-one mapping, then there exists a unique minimum inaccuracy estimate of* θ *for* $(x_1, \ldots, x_n)$ *given by the value* $\theta^*(x_1, \ldots, x_n) \in \Theta$ *such that*

$$\mathbf{h}(\theta^*(x_1, \ldots, x_n)) = \int_{X^n} \mu_{|(x_1,\ldots,x_n)|}(x_1,\ldots,x_n)\, T(x_1,\ldots,x_n)\, d\lambda(x_1) \ldots d\lambda(x_n)$$ ●

Remark 3.2: It is worth emphasizing that when we use the minimum inaccuracy principle the presence of fuzziness in the experimental observations does not necessarily entail an increase in the mean square error with respect to the maximum likelihood one since, according to the preceding theorem, the minimum inaccuracy principle usually supplies estimates consisting in "average" maximum likelihood estimates based on exact data.

Besides, the minimum inaccuracy estimates satisfy the following invariance property:

Theorem 3.2. *Let* $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, *be an experiment where* $\{P_\theta, \theta \in \Theta\}$ *is a parametric family of probability measures which is dominated by the counting or the Lebesgue measure* λ. *Let* $g : \Theta \to \Omega$ *be a one-to-one mapping,* $\Omega = g(\Theta)$ *a set in a euclidean space. If* $\theta^*(x_1, \ldots, x_n)$ *is a minimum inaccuracy estimate of* θ *for* $(x_1, \ldots, x_n)$, *then* $\omega^*(x_1,\ldots,x_n) = g(\theta^*(x_1, \ldots, x_n))$ *is a minimum inaccuracy estimate of* $\omega = g(\theta)$ *in* Ω *for* $(x_1, \ldots, x_n)$. ●

In previous papers (1984a, 1988b) several estimates of the parameters in the usual distributions are determined by means of the minimum inaccuracy principle.

We are now going to prove that under certain natural assumptions a minimum inaccuracy estimate of a parameter for a sample fuzzy information provides us with an approximation of that in Definition 2.2.

Let $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, be an experiment where $\{P_\theta, \theta \in \Theta\}$ is a parametric family of probability measures which is dominated by the counting or the Lebesgue measure λ. Let $\mathcal{X}$ be a f.i.s. associated with $\mathbf{X}$, with r = cardinality of $\mathcal{X}$. From now on, we assume that

i*) If $\{P_\theta, \theta \in \Theta\}$ is dominated by the counting measure, there exist r discrete sets S_x (eventually overlapping) such that $\cup S_x = X$ $(x \in \mathcal{X})$ and with $\mu_x(x) > 0$ if $x \in S_x$, $= 0$ otherwise (that is, S_x = support set of x).

If $\{P_\theta, \theta \in \Theta\}$ is dominated by the Lebesgue measure, there exist r intervals S_x (eventually overlapping) such that $\cup S_x = X$ $(x \in \mathcal{X})$ and with $\mu_x(x) > 0$ if $x \in S_x$, $= 0$ otherwise (that is, S_x = support set of x), μ_x being a continuous membership function on X.

ii*) We agree that $\underline{\lambda}(x_1, \ldots, x_n) = \prod_{x \in \mathcal{X}} [\underline{\lambda}(x)]^{\nu(x)} < \infty$ and $\Im_\theta = -\sum_{x \in \mathcal{X}} \nu(x)\underline{\psi}_\theta(x) < \infty$, provided that $\nu(x) = 0$ implies that $[\underline{\lambda}(x)]^{\nu(x)} = 1$ and $\nu(x)\underline{\psi}_\theta(x) = 0$, whatever $\underline{\lambda}(x)$ and $\underline{\psi}_\theta(x) = \int_X \mu_{|x|}(x) \log f(x;\theta)\, d\lambda(x)$ (finite or infinite) may be. f being the density associated with P_θ with respect to λ.

Under such assumptions the maximum likelihood estimate of θ for the sample fuzzy information $(x_1, ..., x_n)$ as intended in Definition 2.2 could be obtained by minimizing with respect to θ the expression $-\log \mathcal{L}_\theta = -\sum_{x \in \mathcal{X}} v(x) \log \int_{S_x} \mu_{|x|}(x)\, f(x;\theta)\, d\lambda(x)$, whereas the minimum inaccuracy estimate of θ for the sample fuzzy information $(x_1, ..., x_n)$ is obtained by minimizing with respect to θ the expression $\mathfrak{I}_\theta = -\sum_{x \in \mathcal{X}} v(x) \int_{S_x} \mu_{|x|}(x) \log f(x;\theta)\, d\lambda(x)$. Consequently, the application of the Weighted Mean Value Theorem of the Integral Calculus to the integrals in the preceding expressions entails that the first estimate is a value minimizing with respect to θ the function $\sum v(x) \log \bar{f}(\xi_x(\theta); \theta)$ and the second one is a value minimizing with respect to θ the function $\sum v(x) \log \bar{f}(\eta_x(\theta); \theta)$ (where $\xi_x(\theta)$ and $\eta_x(\theta)$ are points in the minimum interval containing S_x, and $\bar{f}(\cdot; \theta)$ is a continuous nonnegative function extending almost surely the density $f(\cdot; \theta)$ associated with P_θ with respect to the measure λ).

The preceding analogy makes evident the approximation between the solutions in both methods under some plausible requirements concerning the membership function μ_x and the support sets S_x. This approximation is formalized in the following theorem:

<u>Theorem 3.3</u>. *Let* $\mathbf{X} = (X, \beta_X, P_\theta)$, $\theta \in \Theta$, *be an experiment where* $\{P_\theta, \theta \in \Theta\}$ *is a parametric family of probability measures which is dominated by the counting or the Lebesgue measure* λ. *Assume that conditions* i*) *and* ii*) *are satisfied. Moreover, suppose that*

iii*) *If the sample space* X *is infinite, the* n *observations from* $\mathbf{X}$ *are classified so that the fuzzy events* x *whose support sets* S_x *are infinite extreme discrete sets or intervals (that is, sets whose infimum or supremum are infinite) are not associated with any observed values.*

iv*) S_x *are constructed so that the greater the observed frequency of* x *the smaller* $\lambda(S_x)$.

Then, under the assumptions i)-iv) (Theorem 3.1), *an approximate solution of the maximum likelihood principle for a sample fuzzy information may be obtained by applying the minimum inaccuracy principle to the same sample fuzzy information.* ●

On the basis of the preceding result we are going to state a theorem enclosing a result leading us to an operative approximation to the *extended likelihood ratio test.*

<u>Theorem 3.4</u>. *Let* $\mathcal{X}^{(n)}$ *be a fuzzy random sample of size* n *from the available f.i.s.* $\mathcal{X}$ *on the random experiment* $\mathbf{X}$, *and assume that the parametric probability measure on* $\mathbf{X}$ *determines a parametric distribution function that is regular in all of its first and second* θ*-derivatives* ($\theta = (\theta_1,...,\theta_k)$ *being a k-dimensional parameter*). *Under the assumptions* i)-iv) *and* i*)-iv*), *and the hypothesis* $H_0 : \theta \in \Theta_0 = \{(\theta_1,...,\theta_k) \in \Theta \mid \theta_1 = \theta_1^\circ, ..., \theta_s = \theta_s^\circ\}$, $(s \leq k)$, *the statistic* $-2 \log \Lambda_n^*$, *where* Λ_n^* *is the statistic based on* $\mathcal{X}^{(n)}$ *such that*

$$\Lambda_n^* = \prod_{x \in \mathcal{X}} [\mathcal{P}_{\theta_0^*}(x) / \mathcal{P}_{\theta^*}(x)]^{\nu(x)}$$

(θ_0^* = minimum inaccuracy estimate of θ in Θ_0 for the obtained sample fuzzy information, θ^* = minimum inaccuracy estimate of θ in Θ for the obtained sample fuzzy information, $\nu(x)$ = observed frequency of x in the obtained sample fuzzy information), has approximately an asymptotic chi-square distribution with s *degrees of freedom.* ●

Remark 3.3: In order to confirm that the orthogonality condition assumed for the set of all available fuzzy observations from the experiment does not mean loss of generality, it is interesting to point out that both, the statistic Λ_n (or Λ_n^*) and the maximum likelihood principle (or the minimum inaccuracy principle), are obviously scale invariant with respect to the membership function of the sample fuzzy information, and to obtain the inference in the extended (or approximated extended) likelihood ratio test only requires in practice to know the considered sample fuzzy information. Consequently, the above scale invariance makes "equivalent" in this test two sample fuzzy observations whose membership functions only differ in a multiplicative constant. Thus, if we knew the set of all available sample fuzzy observations we could easily construct a set containing equivalent sample fuzzy events for all of them and additionally determining an orthogonal system.

Remark 3.4: Clearly, whenever the null hypothesis we wish to test is simple, we obtain that $-2 \log \Lambda_n \geq -2 \log \Lambda_n^*$, so that the rejection of the null hypothesis at a given significance level by means of the approximated extension guarantees the rejection at that level by means of the immediate extension.

4. ILLUSTRATIVE EXAMPLES

We now examine several examples in which the shape of the likelihood function in Definition 2.1 determines a function so that to solve the likelihood equation becomes almost impossible.

Example 4.1: 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 distinguish with sharpness the presence of a microorganism and, consequently, he cannot establish the exact number of microorganisms per square centimeter (that he previously knows has a geometric distribution), but rather he can only perceive one of the following observations: x_1 = "a very small number of microorganisms are found", x_2 = "approximately a number of microorganisms from 5 to 8 are found", x_3 = "a great number of microorganisms are found", that the investigator describes by means of the membership function in Figure 2. (Obviously, we may easily construct a f.i.s. $\mathcal{X} = \{ x_1, x_2, x_3, x_4 \}$ on X = { nonnegative integers }.

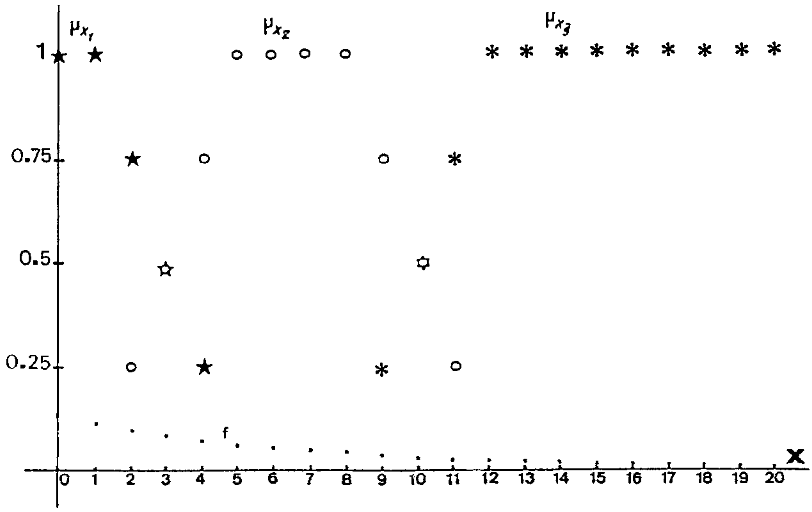

Fig. 2. Membership functions of "a very small number of microorganisms are found" (x_1 , ⋆), "approximately 5 to 8 are found" (x_2 , o), " a great number are found" (x_3 , *), and Geometric probability function (f, •).

In order to test the hypothesis H_0: $p = p_0 = 1/8$, (p = probability of success) the investigator analyzes n=90 slides which provide the observed frequencies $\nu(x_1) = 35$, $\nu(x_2) = 34$, $\nu(x_3) = 21$, $\nu(x_4) = 0$.

According to Theorems 3.1 and 3.2 the minimum inaccuracy estimate of the probability p of success is given by

$$p^* = [1 + \Sigma_j \, \nu(x_j) \, \Sigma_{x \in \mathbb{N}} \, x \, \mu_{|x_j|}(x) \, /90]^{-1} = 0.132353$$

Consequently,

$$\mathcal{P}_{p_0}(x_1) = 0.366341, \; \mathcal{P}_{p_0}(x_2) = 0.384821, \; \mathcal{P}_{p_0}(x_3) = 0.188277,$$

$$\mathcal{P}_{p^*}(x_1) = 0.383891, \; \mathcal{P}_{p^*}(x_2) = 0.388042, \; \mathcal{P}_{p^*}(x_3) = 0.177343,$$

whence $-2 \log \Lambda_n^* = 1.330 < \mathbf{c} = 1.64 = 20\ \%$ point of the chi-square distribution with s=k=1 degree of freedom and, hence, the null hypothesis could be accepted at significance levels lower than $\alpha = 0.20$.

Example 4.2: The observation of the time of attention **X** (in minutes) to a game in four-year old children may be considered as an experiment having exponential distribution and providing imprecise information, since the loss of interest in a game does not usually happen in an instantaneous way. Assume that the mechanism adopted to measure the time of attention only permits us to distinguish the following fuzzy observations: t_1 = "approximately less than 10 minutes", t_j = "approximately 10 j minutes" (j=2,...,8), t_9 = "approximately more than 90 minutes", that we express by means of the membership functions in Figure 3.

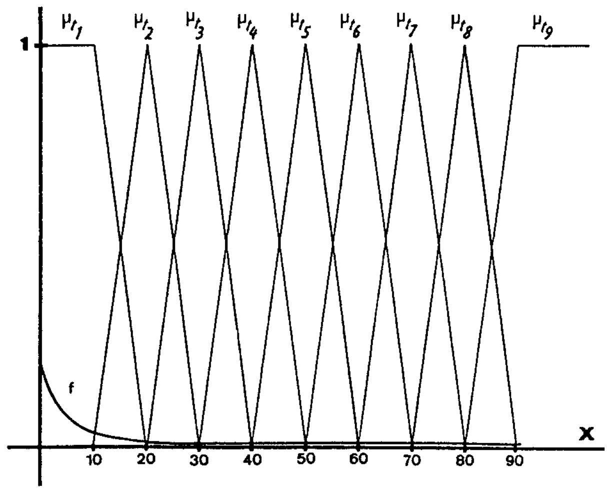

Fig. 3. Membership functions of t_1 = "approximately less than 10 minutes", t_i = "approximately 10 i minutes" (j=2,...,8), t_9 = "approximately more than 90 minutes" and exponential density function with mean 20.

In order to test the null hypothesis H_0: $\mu = \mu_0 = 20$ (μ = population mean of $\mathbf{X}$), a sample of n=600 four-year children have been examined and the following are the observed frequencies:

$\nu(t_1) = 330$, $\nu(t_2) = 119$, $\nu(t_3) = 72$, $\nu(t_4) = 40$, $\nu(t_5) = 19$, $\nu(t_6) = 13$, $\nu(t_7) = 5$, $\nu(t_8) = 2$

Theorems 3.1 and 3.2 indicated that the minimum inaccuracy estimate of μ is given by

$$\mu^* = 600 \, [\textstyle\sum_j \nu(t_j) \int_{\mathbb{R}^+} x \, \mu_{|t_j|}(x) \,]^{-1} = 0.132353$$

Consequently,

$$\mathcal{P}_{\mu_0}(t_1) = 0.587424, \quad \mathcal{P}_{\mu_0}(t_2) = 0.186195, \quad \mathcal{P}_{\mu_0}(t_3) = 0.102165,$$
$$\mathcal{P}_{\mu_0}(t_4) = 0.056058, \quad \mathcal{P}_{\mu_0}(t_5) = 0.030759, \quad \mathcal{P}_{\mu_0}(t_6) = 0.016878,$$
$$\mathcal{P}_{\mu_0}(t_7) = 0.009261, \qquad \mathcal{P}_{\mu_0}(t_8) = 0.005081,$$

$$\mathcal{P}_{\mu^*}(t_1) = 0.521785, \quad \mathcal{P}_{\mu^*}(t_2) = 0.187783, \quad \mathcal{P}_{\mu^*}(t_3) = 0.114045,$$
$$\mathcal{P}_{\mu^*}(t_4) = 0.069263, \quad \mathcal{P}_{\mu^*}(t_5) = 0.042065, \quad \mathcal{P}_{\mu^*}(t_6) = 0.025547,$$
$$\mathcal{P}_{\mu^*}(t_7) = 0.015515, \quad \mathcal{P}_{\mu^*}(t_8) = 0.009428,$$

whence $-2 \log \Lambda_n^* = 12.682 > \mathbf{c} = 7.88$ = 0.5 % point of the chi-square distribution with 1 degree of freedom and, hence, the null hypothesis must be rejected even at significance levels definitively lower than $\alpha = 0.005$.

5. CONCLUDING REMARKS

The results in this paper may be complemented by the analysis of the asymptotic distribution of the statistics in Theorems 2.1 and 3.4 following ideas similar to those developed by Wald (1941a, 1941b), and Wilks (1938a), and those carried out by Cramér (1946), Fisher (1924), Chernoff and Lehmann (1954), for the chi-square test for goodness of fit. In addition, one can extend in an analogous way the likelihood ratio test for goodness of fit.

Finally, the results we have developed in this paper could immediately be particularized to the special case in which the fuzzy information reduces to grouped data (see, Gil and Corral, 1987b).

ACKNOWLEDGEMENT

The authors are grateful for the helpful referee's criticisms.

REFERENCES

Casals, M.R., Gil, M.A. and Gil, P. (1986a). On the use of Zadeh's Probabilistic Definition for Testing Statistical Hypotheses from Fuzzy Information. *Fuzzy Sets and Systems*, 20, 175-190.

Casals, M.R., Gil, M.A. and Gil, P. (1986b). The Fuzzy Decision Problem: an approach to the Problem of Testing Statistical Hypotheses with Fuzzy Information. *European J. Oper. Res.*, 27, 371-382.

Chernoff, H. and Lehmann, E.L. (1954). The use of the Maximum Likelihood Estimates in χ^2 test for Goodness of Fit. *Ann. Math . Stat.*, 25, 579-586.

Corral, N. and Gil, M.A. (1984a). The Minimum Innaccuracy Fuzzy Estimation: an extension of the Maximum Likelihood Principle. *Stochastica*, VIII, 63-81.

Cramer, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press. Princeton.

Fisher, R.A. (1924). The condition under which χ^2 measures the discrepancy between observation and hypothesis. *J. Royal Statist. Soc.*, 87, 442-450.

Gil, M.A. (1987a). Fuzziness and Loss of Information in Statistical problems. *IEEE Trans. Systems Man Cybernet.* , 17, 6, 1016-1025.

Gil, M.A. (1988a). Probabilistic-Possibilistic Approach to some Statistical Problems with Fuzzy Experimental Observations. In *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision-Making* (J. Kacprzyk and M. Fedrizzi, Eds.). Springer-Verlag. Berlin, (accepted for publication).

Gil, M.A. and Corral, N. (1987b). The Minimum Inaccuracy Principle in Estimating Population Parameters from Grouped Data. *Kybernetes*, 16, 43-49.

Gil, M.A., Corral, N. and Gil, P. (1985a). The Fuzzy Decision Problem: an approach to the Point Estimation Problem with Fuzzy Information. *European J. Oper. Res.*, 22, 26-34.

Gil, M.A., Corral, N. and Gil, P. (1988b). The Minimum Inaccuracy Estimates in χ^2 tests for Goodness of Fit with fuzzy observations. *J. Statist. Plan. Inf.*, 18 (accepted for publication).

Gil, M.A., López, M.T. and Gil, P. (1984b). Comparison between Fuzzy Information Systems. *Kybernetes*, 13, 245-251.

Gil, M.A., López, M.T. and Gil, P. (1985b). Quantity of Information; Comparison between Information Systems: 1. Nonfuzzy States, 2. Fuzzy States. *Fuzzy Sets and Systems*, 15, 65-78, 129-145.

Kerridge, D.F. (1961). Inaccuracy and Inference. *J. Royal Stat. Soc.* , *Ser. B*, 23, 184-194.

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

Okuda, T., Tanaka, H. and Asai, K. (1978). A Formulation of Fuzzy Decision Problems with Fuzzy Information, using Probability Measures of Fuzzy Events. *Inform. Contr.*, 38, 135-147.

Rathie, P.N. and Kannappan, Pl. (1973). An Inaccuracy Function of type β. *Ann. Inst. Statis. Math.*, 25, 205-214.

Saaty, T.L. (1974). Measuring the Fuzziness of Sets. *J. Cybernet.*, 4, 4, 53-61.

Tanaka, H. , Okuda, T. and Asai, K. (1979). Fuzzy Information and Decision in Statistical Model. In *Advances in Fuzzy Sets Theory and Applications*. North-Holland. Amsterdam, pp. 303-320.

Wald, A. (1941a). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Stat.*, 12, 1-19.

Wald, A. (1941b). Some examples of asymptotically most powerful tests. *Ann. Math. Stat.*, 12, 396-408.

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio test for testing composite hypotheses. *Ann. Math. Stat.*, 9, 60-62.

Zadeh, L.A. (1965). Fuzzy Sets. *Inform. Contr.*, 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.

María Angeles Gil and María Rosa Casals

Departamento de Matemáticas
Universidad de Oviedo
33071 Oviedo, SPAIN