International Journal of Approximate Reasoning 52 (2011) 240–260

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Mark-recapture techniques in statistical tests for imprecise data

Inés Couso [a,*], Luciano Sánchez [b]

[a] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Gijon, Asturias, Spain
[b] Departamento de Informática, Universidad de Oviedo, Gijon, Asturias, Spain

ARTICLE INFO

Article history:
Received 10 January 2008
Received in revised form 30 July 2010
Accepted 30 July 2010
Available online 15 August 2010

Keywords:
Hypothesis testing
Imprecise data analysis
Multi-valued mappings
Probability bounds
Multi-valued test function

ABSTRACT

We aim to construct suitable tests when we have imprecise information about a sample. More specifically, we assume that we get a collection of n sets of values, each one characterizing an imprecise measurement. Each set specifies where the true sample value is (and where it is not) with full confidence, but it does not provide any additional information.

Our main objectives are twofold: first we will review different kinds of tests in the literature about inferential statistics with random sets and discuss the approach that best suits our definition of imprecise data. Secondly, we will show that we can take advantage from mark and recapture techniques to improve the accuracy of our decisions. These techniques will be specially important when the population is small enough (with respect to the sample size) that recaptures are common. They also seem to be useful when resampling techniques are involved in the decision process.

1. Introduction

The modern theory of random sets was initiated in the seventies, independently by David Kendall [29] and Georges Matheron [36] and its rapid development is mainly due to the following applications: stochastic geometry and representation of imprecise random measurements. On the one hand, stochastic geometry is concerned with random geometric structures, ranging from simple points or line segments to arbitrary closed sets. Stochastic geometry techniques can be applied in a wide range of fields, such as image analysis, telecommunication networks, forestry or environmental research. On the other hand, random sets seem to be the natural representation for imprecisely observed random quantities, when we can only determine a set of values that contains each outcome. Due to the imprecision of devices, we can only say that the outcome belongs to a more or less precise set, but we cannot establish a probability distribution within it indicating different grades of belief for different zones of the set. In this paper, we will focus on this particular application of random sets.

Specifically, we will deal with parametric hypotheses testing. So we first need to clarify the notions of "parameter" and "sample statistic" in our context. From now on, and for the sake of simplicity, we will restrict ourselves to random sets defined on the real line, and to the concepts of mean and variance. We can find in random sets literature at least two alternative procedures to extend those notions.

(A) The first option consists in considering the random set as a "random object" (a measurable mapping in a classical sense). So, it induces a probability measure on a σ-algebra defined over a family of subsets of the real line. Each parameter is calculated as a function of its induced probability measure. Debreu expectation [21] and Körner variance [31] are examples of this approach. Following the same approach, the sample mean and variance are also calculated as functions of the empirical distribution of the random set.

* Corresponding author.
E-mail addresses: couso@uniovi.es (I. Couso), luciano@uniovi.es (L. Sánchez).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.07.009

(B) The second one regards the random set as a collection of random variables (the class of its measurable selections). The parameter is calculated as the set of "admissible" values. Aumann expectation [2] and Kruse variance [33], for instance, fit this formulation. The sample mean and variance are calculated in a similar way, as sets of admissible values for the mean and variance of the underlying imprecisely observed data.
Even when it is defined as a set of admissible values, Aumann expectation is sometimes considered in the literature within the first approach (option (A)), because, under some conditions about the non-atomicity of the initial space, it can be written as a function of the probability distribution of the random set.

Regarding the construction of tests for set-valued data, we also find two different approaches in the literature, each one in accordance with the above alternatives.

(1) In the first approach, the null and the alternative hypotheses concern the "value" of a particular parameter of the random set. Standard binary (always conclusive) tests are proposed to reach a decision (see [6,32,43], for instance). This approach is usually combined with the first option (option (A)), but not always (see [8], for instance).
(2) In the second approach, the random set is understood as the imprecise perception of a standard random variable. Hypotheses about a certain parameter of this random variable are proposed. Thus, imprecision in sample data is translated to tests functions, so they may be inconclusive (see [24] for a detailed discussion about this approach). This approach is in accordance with option (B).

As far as we know, there is no detailed comparison between the above alternatives in the literature. The choice of one or another will depend on the interpretation of the data set. We will get into details in Sections 3 and 4: In Section 3, we will discuss the difference between options (A) and (B). Example 3.1 will reflect a situation where option (A) is the most appropriate, whereas Example 3.2 illustrates option (B). We conclude that we must choose option B, because our random set represents an imprecisely perceived random variable. Hence, our incomplete knowledge about a particular parameter or about a sample statistic will be described by means of a set of admissible values. On the other hand, a discussion about the difference between options (1) and (2) will be provided in Section 4: Example 4.1 illustrates a situation where option (1) is the most suitable one, while Example 4.2 illustrates option (2). In our setting, we do not look for a test of a certain parameter of a random set. What we want is to make inferences about the random variable's expectation or variance, based upon the information provided by a sample of set-valued observations, so we need to choose option (2).

We must remark that there is a strong connection between (A) and (1) and between (B) and (2), in the literature, in the sense that those works considering definitions according to approach (A) usually propose decision making methods according to (1). Nevertheless, even in the cases where the set of admissible values for the parameter can be written as a function of the probability distribution induced by the random set, and therefore, even when the distinction between (A) and (B) does not apply, the distinction between (1) and (2) makes sense. This point will be discussed in detail in Example 4.2.

Classical tests associate a p-value to each possible sample. When sample data are set-valued, we can use set-valued analysis to compute bounds on the test probability (p-value). When the resulting bounds are on one side of the α-level, the decision of the hypothesis test is clear. But when the bounds straddle the threshold, the test is inconclusive, since the imprecision in the data prevents us from making a clear determination. But the more precise they are, the more likely we will be able to make a decision. In Section 4 we illustrate these ideas. Afterwards, in Section 5, we postulate that marking and recapture techniques will exploit data in a suitable way in order to increase the probability of reaching a clear decision. These techniques will be especially useful when the population is small enough that recaptures are common. It will be also useful when resampling techniques are used, as we illustrate in Section 6.

2. Preliminaries and notation

Consider a probability space $(\Omega, \mathcal{A}, P)$, and an arbitrary measurable space $(\Omega', \mathcal{A}')$. Let $f{:}\Omega \to \Omega'$ be an $\mathcal{A} - \mathcal{A}'$ measurable function, i.e., a function satisfying the condition:

$$f^{-1}(A) \in \mathcal{A}, \quad \forall A \in \mathcal{A}'.$$

We will denote by $P \circ f^{-1}$ the probability measure it induces on $\mathcal{A}'$, i.e.

$$P \circ f^{-1}(A) := P(f^{-1}(A)), \quad \forall A \in \mathcal{A}',$$

where $f^{-1} : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega)$ is the mapping defined as

$$f^{-1}(A) = \{\omega \in \Omega : f(\omega) \in A\}, \quad \forall A \subseteq \Omega'.$$

Let us now consider a measurable space, $(\Omega, \mathcal{A})$, the usual Borel σ-algebra on $\mathbb{R}^n$, $\beta_{\mathbb{R}^n}$, and the power set of $\mathbb{R}^n$, $\wp(\mathbb{R}^n)$. Let Γ be a multi-valued mapping between Ω and $\mathbb{R}^n$, $\Gamma : \Omega \to \wp(\mathbb{R}^n)$, with non-empty images. Let $A \in \beta_{\mathbb{R}^n}$ an arbitrary Borel set. The *upper inverse* of A is the set:

$$\Gamma^*(A) = \{\omega \in \Omega : \Gamma(\omega) \cap A \neq \emptyset\}.$$

We say that Γ is *strongly measurable* [44] when $\Gamma^*(A) \in \mathcal{A},\ \forall A \in \beta_{\mathbb{R}^n}$.

Given a probability space $(\Omega, \mathcal{A}, P)$, we will say that the multi-valued mapping $\Gamma : \Omega \to \wp(\mathbb{R}^n)$ is a *random set* when it is strongly measurable. Let us now consider, for each $A \in \beta_{\mathbb{R}^n}$, the family of sets:

$$\mathcal{P}_A = \{C \subseteq \mathbb{R}^n : C \cap A \neq \emptyset\}.$$

Let $\sigma_{\mathcal{P}}$ be the σ-algebra generated by the class $\mathcal{P} = \{\mathcal{P}_A : A \in \beta_{\mathbb{R}^n}\}$ on the universe of "elements" $\wp(\mathbb{R}^n)$. We easily observe that Γ is strongly measurable if and only if it is $\mathcal{A} - \sigma_{\mathcal{P}}$ measurable (regarded as a point to point function, where the "points" in the final space are elements of $\wp(\mathbb{R}^n)$). We will denote by $P \circ \Gamma^{-1}$ the probability measure induced by Γ on $\sigma_{\mathcal{P}}$:

$$P \circ \Gamma^{-1}(\mathcal{C}) = P(\{\omega \in \Omega : \Gamma(\omega) \in \mathcal{C}\}), \quad \forall \mathcal{C} \in \sigma_{\mathcal{P}}.$$

We will denote by $S(\Gamma)$ the class of *measurable selections* of Γ:

$$S(\Gamma) = \{X : \Omega \to \mathbb{R}^n : X \quad \text{measurable and} \quad X(\omega) \in \Gamma(\omega), \quad \forall \omega \in \Omega\}.$$

Let us suppose that Γ represents some incomplete knowledge about the outcomes of a (standard) random variable X^*. In other words, all we know about each outcome $X^*(\omega)$ is that it belongs to the set $\Gamma(\omega)$. Then, all we know about X^* is that it is a measurable selection of Γ. Consider an arbitrary parameter associated to the probability distribution of X^*, $\theta(X^*)$. The random set Γ regards the following information about $\theta(X^*)$:

$$\theta(\Gamma) = \{\theta(X) : X \in S(\Gamma)\}, \tag{1}$$

as suggested by Kruse and Meyer [34]. In other words, if we observe $\Gamma(\omega)$ as an incomplete perception of $X^*(\omega)$, for each $\omega \in \Omega$, then all we know about $\theta(X^*)$ is that it belongs to the class $\theta(\Gamma)$. When, in particular, $\theta(X^*)$ represents the Lebesgue expectation, $E(X^*)$, then $\theta(\Gamma)$ is called the *Aumann expectation*[1] [2] and, when it represents the variance, $\theta(\Gamma)$ is called the *Kruse variance* [33].

The probability measure induced by a random set, $P \circ \Gamma^{-1}$ does not determine its class of measurable selections. Thus, we can find two different random sets, Γ_1 and Γ_2, with the same probability distribution, $P \circ \Gamma_1^{-1} = P \circ \Gamma_2^{-1}$, but with different collections of measurable selections, $S(\Gamma_1) \neq S(\Gamma_2)$. (See [17,19,37–40] for detailed discussions.) Furthermore, the classes of values $\theta(\Gamma_1)$ and $\theta(\Gamma_2)$ can be different, for some parameter θ, such as the expectation, the variance, etc.

Let $X^* : \Omega \to \mathbb{R}$ be a random variable representing some numerical characteristic of the individuals in the population Ω. A random sample of size n is a random vector $(X_1^*, \ldots, X_n^*) : \Omega^n \to \mathbb{R}^n$, where each $X_i^* : \Omega^n \to \mathbb{R}$ is defined as follows:

$$X_i^*(\omega_1, \ldots, \omega_n) = X^*(\omega_i), \quad \forall (\omega_1, \ldots, \omega_n) \in \Omega^n.$$

(A collection, of n elements – some of them maybe coincident – is selected from the population. Such a collection of elements is denoted by $(\omega_1, \ldots, \omega_n) \in \Omega^n$. Each X_i^* represents the value of the characteristic for the element selected at random in the ith place.) When, in particular, the product probability is considered over Ω^n (i.e., when the individuals are taken with replacement, and each selection is independent from the others), the random variables $X_1^*, \ldots, X_n^*$ are i.i.d., so they represent a simple random sample.

Let now the multi-valued mapping $\Gamma : \Omega \to \wp(\mathbb{R})$ represent some incomplete knowledge about $X^* : \Omega \to \mathbb{R}$, in the sense that, when we select a particular element ω from the population Ω, we cannot observe the quantity $X^*(\omega)$, but we just know that it belongs to the set $\Gamma(\omega)$. We will represent the imprecise observation of the random sample $(X_1^*, \ldots, X_n^*)$ by means of the multi-valued mapping $\vec{\Gamma} = (\Gamma_1 \times \ldots \times \Gamma_n) : \Omega^n \to \wp(\mathbb{R}^n)$, defined as

$$(\Gamma_1 \times \cdots \times \Gamma_n)(\omega_1, \ldots, \omega_n) = \Gamma_1(\omega_1, \cdots, \omega_n) \times \cdots \times \Gamma_n(\omega_1, \ldots, \omega_n),$$

where

$$\Gamma_i(\omega_1, \ldots, \omega_n) = \Gamma(\omega_i), \quad \forall (\omega_1, \ldots, \omega_n) \in \Omega^n.$$

Summarizing, for a specific sequence of n selections $(\omega_1, \ldots, \omega_n) \in \Omega^n$, all we know about the realization $\vec{x}^* = (x_1^*, \ldots, x_n^*) = (X^*(\omega_1), \ldots, X^*(\omega_n))$ is that it belongs to the cartesian product $\gamma = \gamma_1 \times \cdots \times \gamma_n = \Gamma(\omega_1) \times \cdots \times \Gamma(\omega_n)$.

3. Two different ways to define parameters and sample statistics for random sets

As we have pointed out in Section 1, a random set can be understood either as a family of measurable selections or as a measurable set-valued mapping. Such a measurable function induces a probability measure on the final set. Let us notice that this probability distribution does not determine, in general, the class of measurable selections of a random set, as shown in [17,37–39,42,47]. Thus, we can find two different random sets with the same probability distribution, but with different Aumann expectations and/or different Kruse variances (see [11,17], for instance). Hence a hypothesis test about the "value" of the parameter according to the first formulation would not be informative at all to our purposes. Let us next compare both approaches in more detail.

[1] In fact, the set of all integrable selections, $S^1(\Gamma) \subseteq S(\Gamma)$ instead of the whole class $S(\Gamma)$ is considered here.

3.1. First approach: the random set as a random object

According to the first approach (option (A) described in Section 1), the random set is just considered as a particular measurable mapping within the framework of classical probability theory. Thus, the parameter is calculated as a function of the probability distribution of the random set. The concepts of expectation, variance, etc. are extended by reproducing classical techniques. More specifically, set-valued arithmetic is used to derive a method of construction of the expectation: a limit-based construction analogous to Lebesgue integral definition (using Minkowski sum and scalar product to define the expectation of "simple" random sets) leads to a definition of expectation which is consistent with Bochner integral. This expectation is a subset of the final space and it plays the role of the "average value" of the random set. We can make a parallel construction of the variance: let us consider a particular metric defined over the class of subsets of the final space. In this setting, we can define the variance of a random set as the mean (classical expectation of a random variable) of the squares of the distances from the images of the random set to the (set-valued) expectation. For instance the definitions of Feng et al. [23] and Körner [31] fit this formulation. In this context the variance is a (precise) number that quantifies the degree of dispersion of the images of the random set.

Similarly, the sample mean is calculated by using set-valued arithmetic and the sample variance averages the squared distances (for a particular metric defined for pairs of subsets of the real line) between the sample values and the sample mean.

The outcomes of the experiment considered in the next example are represented, in a natural way, by means of intervals of the real line, so it illustrates this first option.

Example 3.1. In [6], the following collection of 59 intervals is provided:

11.8–17.3	11.9–21.2	9.8–16.0	10.4–16.1	12.2–17.8	9.7–15.4	13.1–18.6	12.0–18.9
8.7–15.0	10.5–15.7	11.3–21.3	14.1–25.6	12.0–17.9	14.1–20.5	10.8–14.7	10.1–19.4
9.9–16.9	11.5–19.6	10.9–17.4	12.6–19.7	9.9–17.2	12.8–21.0	9.9–20.1	11.3–17.6
9.4–14.5	8.8–22.1	11.4–18.6	14.8–20.1	11.3–18.3	14.5–21.0	11.1–19.2	9.4–17.6
12–0-18.0	11.6–20.1	10.2–15.6	10.0–16.1	10.2–16.7	10.3–15.9	15.9–21.4	10.4–16.1
10.2–18.5	13.8–22.1	10.6–16.7	11.1–19.9	8.7–15.2	11.2–16.2	13.0–18.0	12.0–18.8
13.6–20.1	10.3–16.1	9.5–16.6	9.0–17.7	12.5–19.2	9.2–17.3	11.6–16.8	9.7–18.2
8.3–14.0	9.8–15.7	12.7–22.6					

It corresponds to a sample of 59 patients in a hospital, and each interval represents the systolic blood pressure range over a day for each patient. Each interval represents a range of values associated to some patient in a particular day, and it does not represent any incomplete knowledge about a real-valued quantity. The sample mean is calculated in [6] as the closed interval:

$$\bar{\gamma} = \frac{1}{59}(\gamma_1 \oplus \cdots \oplus \gamma_{59}),$$

where $\oplus$ represents the Minkowski sum and $\gamma_i = [m_i - \delta_i, m_i + \delta_i]$ is the interval associated to the ith patient. (In other words, the sample mean is the interval bounded by the arithmetic means of the bounds of the intervals in the sample). On the other hand, the sample variance is calculated as follows:

$$s^2 = \frac{1}{58}\left[d^2(\gamma_1, \bar{\gamma}) + \cdots + d^2(\gamma_{59}, \bar{\gamma})\right],$$

where d is a metric defined for pairs of closed intervals as

$$d\left([m_i - \delta_i, m_i + \delta_i], [\bar{m} - \bar{\delta}, \bar{m} + \bar{\delta}]\right) = \left[(m_i - \bar{m})^2 + \frac{1}{3}(\delta_i - \bar{\delta})^2\right]^{1/2}.$$

In this example, the sample mean, $\bar{\gamma} = [11.19, 18.17]$, represents the "average range" over the 59 patients. The sample variance $s^2 = 1.9$ quantifies the dispersion over the 59 intervals. Several authors (see Feng [23] and Körner [30], for instance) in the literature have proposed other alternative scalar variances, corresponding to different metrics over the class of closed intervals. Each metric would provide a different value for the sample variance, but for all of them the variance is a function of the empirical distribution of the random set (it is determined by the – relative – frequencies of appearance of each interval).

Remark 3.1. Following similar procedures, the expectation and the (population) variance of a random interval can be calculated as functions of the probability measure that it induces on the family of intervals of the real line.

Remark 3.2. In a quite different context, Billard and Diday [5], also calculate the sample mean and variance as functions of the empirical distribution of the random set.

3.2. Second approach: the random set as a family of admissible random variables

According to the second approach, the random set is understood as a family of measurable mappings (the class of its measurable selections). In this context, the parameter is defined as the class of "admissible" values (i.e., the set of values associated to every measurable selection of the random set). According to this approach, Aumann defines the expectation of the random set as the class of expectations of all its integrable selections [2]. Kruse follows the same approach to define the variance [33] of a random set. This approach is related to the concept of "identification region" (see, for instance, [4,26,27,35]) or "identified set" [48,53], terms used in some papers concerning other different incomplete data problems. The identification region is defined in [26] as the set of values of the parameter associated to the family of "feasible distributions". Such family is the set of probabilities consistent with some available empirical information, not necessarily determined by the family of selections of a random set.

In the next example, the random set represents an incompletely perceived random variable and we intend to represent the imprecise knowledge about the parameter, so it illustrates this second alternative. It has been taken from [10]. In that paper a more detailed discussion is given about the notion of variance for random sets and fuzzy random variables.

Example 3.2.

(a) The sample $(\omega_1,\ldots,\omega_4)$ comprises four objects, whose actual weights are $x_1^* = 10.2$, $x_2^* = 10.0$, $x_3^* = 10.4$, $x_4^* = 9.7$. We measure the weights with a digital device that rounds the measure to the nearest integer, and displays the value '10' in all of these cases. Therefore, we get the same interval of values for every object, $\gamma_i = [9.5, 10.5]$, $\forall i = 1,\ldots,4$. The true variance of the four measurements is 0.067. But, we only know the information provided by the four intervals, so all we can say about the variance is that it is bounded by the values 0 and 0.25. (Kruse's variance returns this range of values). On the contrary, the scalar variance (Körner and Feng definitions) returns the misleading value 0.

(b) Let us modify a bit the last situation. Let us suppose that, instead of four objects, we had only selected the first one (suppose that we had a sample of size $n = 1$). Then, Kruse's variance would returns the singleton {0}. (We know with certainty that the variance of a single measurement is 0). So, let us observe that Kruse's variance cannot be written as a function of the empirical distribution of the random set. In both cases (cases (a) and (b) in this example), the (relative) frequency of the interval [9.5,10.5] is 1, but Kruse's variance returns different ranges of values. On the contrary, Körner's and Feng's variances can be written as functions of the empirical distribution. Hence, Körner's and Feng's variances would be useless within this context.

(c) Let us suppose that four objects $\omega_1,\ldots,\omega_4$ weigh the same: $x_1^* = x_2^* = x_3^* = x_4^* = 9.8$ g. Let us also suppose that, for some reason, the weight of the fourth object was measured with imprecision, and we only know that it is between the values 9.5 and 10.5. Our knowledge about the four measurements are given by $\gamma_1 = \{9.8\}$, $\gamma_2 = \{9.8\}$, $\gamma_3 = \{9.8\}$ and $\gamma_4 = [9.5, 10.5]$. The true variance in the sample is 0 and Kruse's variance produces the interval [0,0.092]. According to the above incomplete information, Kruse's interval represents all our knowledge about the true variance. On the other hand, the scalar variance (Körner and Feng definitions) assigns a strictly positive value to it. We conclude that the scalar variance is neither an upper bound of the actual value of the variance of the underlying sample (see case (a) in this example), nor a lower bound (see case (c)).

(d) Now, suppose that in cases (a) and (b) we are censuring the whole population (the respective population sizes are 4 and 1). Then, the empirical distribution turns into the probability distribution. From (a) and (b), we conclude that Kruse's variance is not a function of the probability distribution of the random set (in both cases, the random sets take the "value" [9.5,10.5] with probability 1, but Kruse's variance does not return the same interval. So, when we aim to describe all the available (incomplete) information about a parameter, the probability distribution induced by the random set does not provide enough information.

It has been proved by Castaldo et al. [7, Corollary 3.4], that the weak closure of the family of probability distributions induced by the measurable selections of any compact-valued random set defined on a non-atomic space, and taking values on a Polish space, is determined by the probability distribution of the random set (considered as a measurable mapping). This result has been independently proved and also extended to other situations by Miranda et al. in [37, Corollary 2] and [39, Corollary 4.8]. Those situations cover the cases where the images of the random set are open or closed sets in a Polish space, among others. According to these results, the differences between approaches A and B are irrelevant when the initial space is non-atomic, and the images of the random set are open or closed subsets of the real line. In such cases, the set of admissible values for a specific parameter can be essentially written as a function of the probability distribution induced by the random set. Let us furthermore notice that the non-atomic case covers some important specific situations, as the case where the probability distribution of the original random variable is assumed to be absolutely continuous (w.r.t. Lebesgue measure).

So far, we have not approached the problem of making inferences from set-valued data samples. As we have suggested in the introduction, we can distinguish in the literature two different types of methods to construct hypothesis tests for fuzzy/crisp-set-valued data. The first one leads to classical tests about fuzzy/crisp-set-valued parameters, since the second method ([24,45,25]) leads to fuzzy/crisp-set-valued tests about precise parameters. This second approach is closely related to the theory of imprecise probabilities. In the following subsection, we will compare both approaches and justify why the second approach is the one we will choose in our context.

4. Two different ways to make inferences from set-valued data

4.1. Inferences about the probability distribution induced by the random set

Körner suggests in [30] a general procedure for testing hypothesis about fuzzy parameters associated to fuzzy random variables. It can be particularized to the case of random sets. It essentially consist in treating the random set as a "random object" that induces a probability measure on the power set of the real line (assigning a probability value to each class of subsets of the real line), and making inferences about a particular parameter (possibly set-valued), θ, associated to such probability distribution. A metric, d, between pairs of subsets of the real line and a consistent estimator, $\hat{\theta}$, of θ (consistent with respect to the distance d) are considered. As Körner suggests, we are inclined to reject H_0: $\theta = \theta_0$ against H_1: $\theta \neq \theta_0$ when the distance $d(\hat{\theta}, \theta_0)$ is too large. Since d is a standard metric (it returns precise values), $d(\hat{\theta}, \theta)$ is (under some measurability restrictions) a standard random variable. When the exact or the asymptotic probability distribution of this random variable is known, a standard α-test can be easily derived. In fact, we should reject the null hypothesis when

$$d(\hat{\theta}, \theta_0) > t_{1-\alpha},$$

where t_q is the q-quantile of $d(\hat{\theta}, \theta)$ and α is the type I error. Thus:

$$P\Big(d(\hat{\theta}, \theta_0) > t_{1-\alpha} | H_0\Big) \leqslant \alpha.$$

We can find in the literature different papers based on these ideas (see [6,32,43], for instance). Each of these papers determines the probability distribution of the statistic $d(\hat{\theta}, \theta_0)$ for a particular distance d and a particular parameter θ. Let us illustrate this procedure with an example.

Example 4.1. The following table has been taken from [43], and it corresponds to a random sample of 59 patients from the Nephrology Unit of a hospital. Each interval $\gamma_i = [m_i - \delta_i, m_i + \delta_i]$ represents the range for the pulse rate during a day for a particular patient. The interval [60,100] is assumed to be the adequate pulse rate fluctuation in a healthy adult. We aim to test whether the range for the pulse rate of the patients in that unit is similar to the range of healthy people. Let Γ denote the random interval corresponding to the pulse rate range of patients selected at random from the Nephrology Unit. We want to test the hypothesis $H_0 : E(\Gamma) = [60,100]$ against $H_1 : E(\Gamma) \neq [60,100]$. In [43] it is proposed to reject H_0 when the following inequality holds:

$$\pi t_m^2 + (1-\pi) t_\delta^2 > t_{1-\alpha},$$

where

$$t_m^2 = \frac{\left(\frac{\sum_{i=1}^{59} m_i}{59} - 80\right)^2}{s_m^2}, \quad t_\delta^2 = \frac{\left(\frac{\sum_{i=1}^{59} \delta_i}{59} - 20\right)^2}{s_\delta^2}, \quad s_m^2 = \frac{\sum_{i=1}^{59}\left(m_i - \frac{\sum_{i=1}^{59} m_i}{59}\right)^2}{58}, \quad s_\delta^2 = \frac{\sum_{i=1}^{59}\left(\delta_i - \frac{\sum_{i=1}^{59} \delta_i}{59}\right)^2}{58}, \quad \pi = \frac{s_m^2}{s_m^2 + \frac{1}{3} s_\delta^2},$$

and the quantile $t_{1-\alpha}$ is determined using bootstrap techniques.

This test is based on the fact that the equality $E(\Gamma) = [60,100]$ equates to $E(\text{mid}\,(\Gamma)) = 80$ and $E(\text{spr}\,(\Gamma)) = 20$, where $\text{mid}(\Gamma)$ and $\text{spr}(\Gamma)$, respectively, denote the midpoint and the radius of Γ.

The p-value associated to the data set is $p = 0.0003$, so there are enough evidences to reject the hypothesis that the pulse rate range of those patients coincides with that for healthy people.

4.2. Inferences about the probability distribution induced by the underlying random variable

The last method is not useful when the random set represents the imprecise observation about a random variable, and we aim to make inferences about its distribution. In that case, we must follow the procedure detailed in Ferson et al. [24] (option 2, in the introduction of our paper), as we will illustrate in the following example.

Example 4.2. The random sample $\{\omega_1, \ldots, \omega_{100}\}$ comprises a hundred objects, and it has been taken from a large population. Their actual weights (X) are displayed in the following frequency:

x_i	5.5	6.5	6.6	7.6
n_i	27	24	25	24

Let us suppose that we want to test the hypotheses:

$$H_0 : \mu \leqslant 6.5 \quad \text{against} \quad \mu > 6.5, \tag{2}$$

where $\mu = E(X^*)$ represents the expectation of the actual weight. Having into account that the statistic $T_n = \frac{\sqrt{n}(\bar{X}_n^* - E(X^*))}{S_{X^*}}$ is asymptotically $N(0,1)$, we should reject the null hypothesis if $\frac{10(\bar{x}^*_{100} - 6.5)}{s^*_x} > t_{1-\alpha}$. This statistic takes the value 0.253, so the p-value is approximately 0.4. Hence, we have no evidences to reject H_0.

Now suppose that we measure the weights with a digital device that rounds the measure to the nearest integer. Thus we obtain a random sample from an interval-valued random set Γ, which is determined by the following frequency table:

If we can only see the information provided by the second table (with the grouping intervals) we can only calculate an interval of bounds for the statistic an the p-value. The bounds of the statistic are -0.42 and 16.44, respectively, so the respective bounds for the p-value are approximately 0 and 0.6628. So, due to the high imprecision in the new data set, our test is inconclusive.

In this situation, a test about the Aumann expectation of Γ would not be appropriate to draw conclusions about H_0. In general, a null hypothesis about the Aumann expectation $E(\Gamma) = [E(L), E(U))$ would refer to the pair of expectations of the observable random variables, $E(L)$ and $E(U)$, where L and R, respectively denote $L = \inf\Gamma$ and $R = \sup\Gamma$. Hence, it may involve some concerns about the degree or imprecision of the observations which are unrelated to the expectation of the underlying random variable, $E(X^*)$. For instance, the null hypothesis H_0' in the test:

$$H_0' : E(\Gamma) = [0, 6.5] \quad \text{against} \quad H_1' : E(\Gamma) \neq [0, 6.5] \tag{3}$$

would represent the assertion: "the average (interval-valued expectation) of the imprecise observations of X^* is [0,6.5]". Thus, $H_0' : E(\Gamma) = [0, 6.5]$ implies that the expected degree of imprecision of the incomplete data is equal to 6.5, which is not implied by the initial null hypothesis H_0. The data set in Table 1 provides enough information to reject H_0' (the specific calculations are omitted), while it is uninformative about H_0. Another candidate concerning the Aumann expectation of Γ could be, for instance the following test:

$$H_0'' : [m, 6.5] \subseteq E(\Gamma) \quad \text{against} \quad H_1'' : [m, 6.5] \not\subseteq E(\Gamma),$$

for some specific choice of m. This test is not appropriate, either, since the null hypothesis H_0'' also involves a specific assumption about the degree of imprecision of the observations (the expected imprecision is less than or equal to $6.5 - m$), which is not covered by the initial hypothesis. The interval data set in Table 1 does not provide evidence to reject H_0''. But we could get some other different data set providing evidence to reject H_0'', and not to reject the initial hypothesis H_0. Furthermore, even for the specific case of the actual data set (Table 1), the conclusions concerning H_0 and H_0'' are not the same: on the one hand, the test concerning H_0 is inconclusive, as we cannot determine whether the actual sample of X^* belongs to the rejection region or not, on the basis of the interval data set provided in Table 1. In other words, the high degree of imprecision in the data set prevents us from taking a decision (reject/no reject) about H_0. On the other hand, the test concerning H_0'' is conclusive, for this (and for any other) data set. The distinction between "not rejecting" and "not concluding" is of interest in practice.

4.2.1. How do we make inferences from imprecise data?

Let $X^* : \Omega \to \mathbb{R}$ be a random variable and let us state the hypothesis:

$$H_0 : \theta(X^*) \in \Theta_0 \quad \text{against} \quad H_1 : \theta(X^*) \in \Theta_1,$$

where $\theta(X^*)$ is a parameter that depends on the probability distribution induced by X^*. Let $\varphi : \mathbb{R}^n \to \{0, 1\}$ be a non-randomized test that represents the decision rule that will lead to a decision to accept or reject the null hypothesis. The critical region associated to φ is

$$C = \{\vec{x} \in \mathbb{R}^n : \varphi(\vec{x}) = 1\}.$$

The mapping φ is said to be a test with level of significance α, $0 \leqslant \alpha \leqslant 1$, when

$$E_\theta(\varphi) = P_\theta(C) \leqslant \alpha, \quad \forall \theta \in \Theta_0.$$

Furthermore, let us suppose that we have an α-test, φ_α, for each $\alpha \in (0, 1]$ with critical region C_α. Let us suppose that those critical regions are one included in another, as usual, i.e.,

$$\left[\alpha_1 \leqslant \alpha_2 \Rightarrow C_{\alpha_1} \subseteq C_{\alpha_2}\right].$$

The critical level or the p-value associated to this family of tests is the mapping $p : \mathbb{R}^n \to [0, 1]$ defined as

$$p(\vec{x}) = \sup\{\alpha \in [0, 1] : \vec{x} \in C_\alpha\}.$$

Let us now consider, for a fixed $\alpha \in [0, 1]$, the test $\varphi'_\alpha : \mathbb{R}^n \to \{0, 1\}$ defined as follows:

Table 1
Interval data set.

γ_i	[5.5, 6.5)	[6.5, 7.5)	[7.5, 8.5)
n_i	27	49	24

$$\varphi'_\alpha(\vec{x}) = \begin{cases} 1 & \text{if} \quad p(\vec{x}) \in [0,\alpha], \\ 0 & \text{otherwise.} \end{cases}$$

It is well know in statistics that φ'_α is an α-test and, under some continuity conditions, it coincides with φ_α.

It is also known in standard statistics that the construction of parametric tests can be based upon the construction of confidence regions. Let us consider a confidence region for the parameter $\theta(X^*)$ at a confidence level $1-\alpha$, $CR : \mathbb{R}^n \to \wp(\mathbb{R})$. It is well known that the mapping $\varphi''_\alpha : \mathbb{R}^n \to \{0,1\}$ defined as follows:

$$\varphi''_\alpha(\vec{x}) = \begin{cases} 1 & \text{if} \quad CR(\vec{x}) \cap \Theta_0 = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

is an α-test of $H_0 : \theta(X^*) \in \Theta_0$ against $H_1 : \theta(X^*) \in \Theta_1$.

Now, assume that we want to make a decision on the basis of a sample of n imprecise measurements. First, let us recall what kind of "imprecise measurements" do we have. Each individual measurement is characterized by an interval (or, more generally, by a set) of values, $\gamma_i \subseteq \mathbb{R}$. The set γ_i specifies where the value x_i^* is (and where is not). Thus, all we know about the sample realization $\vec{x}^* = (x_1^*, \ldots, x_n^*)$ is that it belongs to the cartesian product $\gamma = \gamma_1 \times \cdots \times \gamma_n \in \wp(\mathbb{R}^n)$.

Let us consider a particular test $\varphi : \mathbb{R}^n \to \{0,1\}$. We will assign to γ the (set-valued) decision:

$$D_\varphi(\gamma) = \{\varphi(\vec{x}) : \vec{x} \in \gamma\} \subseteq \{0,1\} = \begin{cases} \{1\} & \text{if} \quad \gamma \subseteq C, \\ \{0\} & \text{if} \quad \gamma \subseteq C^c, \\ \{0,1\} & \text{if} \quad \gamma \cap C \neq \emptyset, \quad \text{and} \quad \gamma \cap C^c \neq \emptyset, \end{cases} \tag{4}$$

D_φ represents our decision rule when we observe data with imprecision. According to this "extended α-test", we will:

- Reject H_0, when $D_\varphi(\gamma) = \{1\}$.
- Do not reject H_0 when $D_\varphi(\gamma) = \{0\}$.
- Do not make any decision when $D_\varphi(\gamma) = \{0,1\}$ (in this case, the test is inconclusive).

In particular, for the tests φ' and φ'' above defined, we can consider the multi-valued decisions $D_{\varphi'}$ and $D_{\varphi''}$. We can easily see that they can be alternatively written as follows:

- Alternative expression for $D_{\varphi'}$:

$$D_{\varphi'}(\gamma) = \begin{cases} \{1\} & \text{if} \quad p(\gamma) \subseteq [0,\alpha], \\ \{0\} & \text{if} \quad p(\gamma) \subseteq (\alpha,1], \\ \{0,1\} & \text{if} \quad p(\gamma) \cap [0,\alpha] \neq \emptyset \quad \text{and} \quad p(\gamma) \cap (\alpha,1] \neq \emptyset, \end{cases} \tag{5}$$

where

$$p(\gamma) = \{p(\vec{x}) : \vec{x} \in \gamma\}.$$

Note that $p(\gamma)$ represents our knowledge about the p-value $p(\vec{x}^*)$. I.e., if all we know about $\vec{x}^* = (x_1^*, \ldots, x_n^*)$ is that $x_i^* \in \gamma_i$, $\forall i = 1, \ldots, n$, then all we know about $p(\vec{x}^*)$ is that it belongs to $p(\gamma)$. We can find in the literature some precedents of the above construction in Ferson et al. [24], Filtzmoser and Viertl [25] and Denoeux et al. [22].

- Alternative expression for $D_{\varphi''}$:

$$D_{\varphi''}(\gamma) = \begin{cases} \{1\} & \text{if} \quad \overline{CR}(\gamma) \cap \Theta_0 = \emptyset, \\ \{0\} & \text{if} \quad \underline{CR}(\gamma) \cap \Theta_0 \neq \emptyset, \\ \{0,1\} & \text{otherwise,} \end{cases} \tag{6}$$

where $\underline{CR}(\gamma)$ and $\overline{CR}(\gamma)$ are defined as

$$\underline{CR}(\gamma) = \cap_{\vec{x} \in \gamma} CR(\vec{x}), \quad \overline{CR}(\gamma) = \cup_{\vec{x} \in \gamma} CR(\vec{x}). \tag{7}$$

The three procedures described in Eqs. (4), (5) and (6) are essentially equivalent (if the corresponding standard tests φ, φ' and φ^{Prime} do coincide, then the set-valued test functions D_φ, $D_{\varphi'}$ and $D_{\varphi''}$ do also coincide). Along the rest of the paper, we will use the procedure described in Eq. (5), which is, in general, the most operative. The idea behind the construction of the "outer approximation", $\overline{CR}(\gamma)$, of the incompletely known confidence region $CR(\vec{x}^*)$ is related to some confidence sets proposed in [1,4,27]. We will provide a more detailed comparison in Section 4.2.2. More details about inner and outer approximations of confidence regions can be found in [18].

When using our set-valued sample data, we will start from a specific α-test (obtained from a specific confidence region, for example) and then we will extend it to the set-valued case, by means of Eq. (4). The evaluation of optimal initial standard tests is out of the scope of the paper. Just as a matter of example, we will next illustrate that the UMP test is not necessarily the best choice:

Example 4.3. Let X^* be a random variable with known standard deviation $\sigma = 1$ and unknown expectation. Let us test the null hypothesis:

$$H_0 : E(X) = \mu_0 \quad \text{against} \quad H_1 : E(X) \neq \mu_0,$$

but let us assume that we cannot get a random sample from X^*, as it happened in the last example. Let us assume that we can only observe a sample of size $n = 100$ of random variable L and that we just know that $L \leqslant X \leqslant L + 1$. In other words, we observe a vector $(l_1,\ldots,l_{100})$ and all we know about $(x_1^*,\ldots,x_n^*)$ is that x_i^* belongs to the closed interval $\gamma_i = [l_i, l_i + 1]$, $\forall i = 1,\ldots,100$.

Let us consider two different confidence regions of $E(X^*)$ at the same level of confidence 0.95:

1. $CR_1(\vec{x}) = (\bar{x} - 0.196, \bar{x} + 0.196)$
2. $CR_2(\vec{x}) = (\bar{x} - 0.164, \infty)$

The first one is associated to the UMP test of size 0.05. The second one is also associated to a 0.05-size test, but it is not the UMP. Let us now apply Eq. (6) to derive a set-valued test for our set valued information. Let $\gamma = \gamma_1 \times \cdots \times \gamma_n = [l_1, l_1 + 1] \times \cdots \times [l_n, l_n + 1]$. According to Eq. (7):

$$\underline{CR_1}(\gamma) = \cap_{\vec{x}\in\gamma} CR_1(\vec{x}) = \emptyset,$$
$$\overline{CR_1}(\gamma) = \cup_{\vec{x}\in\gamma} CR_1(\vec{x}) = (\bar{l} - 0.196, \bar{l} + 1.196),$$

and, according to Eq. (6), we can construct the following set-valued test:

$$D_1(\gamma) = \begin{cases} \{1\} & \text{if } \mu_0 \notin (\bar{l} - 0.196, \bar{l} + 1.196) \\ \{0,1\} & \text{otherwise} \end{cases}$$

For the second confidence region we calculate:

$$\underline{CR_2}(\gamma) = \cap_{\vec{x}\in\gamma} CR_2(\vec{x}) = (\bar{l} + 1 - 0.164, \infty),$$
$$\overline{CR_2}(\gamma) = \cap_{\vec{x}\in\gamma} CR_2(\vec{x}) = (\bar{l} - 0.164, \infty),$$

and the set-valued test:

$$D_2(\gamma) = \begin{cases} \{1\} & \text{if } \mu_0 < \bar{l} - 0.164, \\ \{0\} & \text{if } \mu_0 > \bar{l} + 0.836, \\ \{0,1\} & \text{otherwise.} \end{cases}$$

Notice that the first test (D_1) is inconclusive for the family of samples where:

$$\bar{l} - 0.196 < \mu_0 < \bar{l} - 0.164,$$

while the second test (D_2) rejects the null hypothesis. Furthermore, when:

$$\bar{l} + 0.836 < \mu_0 < \bar{l} + 1.196,$$

the first test is also inconclusive and the second test is conclusive (according to D_2, we do not reject H_0 for those samples). On the other hand, the error II probability bounds are more accurate in the first case than in the second case. So, none of those tests seems to be preferable to the other one. And our preference will depend on the particular application.

4.2.2. Relationship with partial identification problems

There is a common rationale behind this work and many others in the "partial identification" literature (see, for instance, [4,26,27,35,48,53]): the "worst case scenario" assumption. According to this principle, we try to exploit all the available information without adding any artificial knowledge. Notwithstanding the overall philosophy is the same, we cannot take advantage from partial identification techniques, because we are focused on a different kind of problem. To illustrate this fact, we will provide a formal comparison with some techniques initially proposed by Imbens and Manski in [27].

As we pointed out in Section 3.2, the definitions of parameters of a random set according to the "set of admissible random variables" approach is closely related to the notion of "identification region" in partial identification problems. In either case, the set of admissible values for the parameter is considered, and each admissible value is associated to a feasible probability distribution. The difference between both definitions lies in the natures of the families of the probability measures being considered. In this paper we are considering the family of probabilities associated to the measurable selections of the random set, which is not necessarily the case in partial identification problems.

The paper by Imbens and Manski [27] marked a new starting point in confidence estimation for partially identified problems: they showed that, in some situations, for a specific asymptotic confidence level, confidence intervals for the parameter

are proper subsets of (same confidence level) intervals for the identification region, the difference in width being related to the difference in critical values for one- and two-sided tests. This is one of the main differences with respect to our proposal, since our "outer approximation", $\overline{CR}(\gamma)$, covers the whole set-valued parameter of the random set. We will illustrate the difference between both approaches with an example:

Example 4.4. Let X^* be a random variable with unknown expectation and known standard deviation, $\sigma = 1$. Let us test the null hypothesis:

$$H_0 : \mu = \mu_0 \quad \text{against } H_1 : \mu \neq \mu_0,$$

where μ denotes the expectation of X^*. Let us suppose that we cannot get a random sample from X^* but we can only observe a sample of size $n = 100$ of another random variable $L = X^* - \alpha$, where α is an unknown constant within the unit interval [0,0.15]. In other words, we observe a vector $(l_1,\ldots,l_{100})$ and all we know about $(x_1^*,\ldots,x_n^*)$ is that $x_i^* = l_i + \alpha,\ \forall i = 1,\ldots,100$. We can easily check that the random interval $(\bar{l} - 0.165, \bar{l} + 0.315)$ (where $\bar{l}$ denotes the sample mean) coincides with $(\bar{x}^* - 0.165 - \alpha, \bar{x}^* + 0.165 + (0.15 - \alpha))$, and it is an approximately 95% confidence interval for μ, for every $\alpha \in [0,1]$. According to it, the test $\varphi : \mathbb{R}^n \to \{0,1\}$ defined as

$$\varphi(\vec{l}) = \begin{cases} 1 & \text{if } \mu_0 \notin (\bar{l} - 0.165, \bar{l} + 0.315), \\ 0 & \text{otherwise}, \end{cases}$$

is a test of size 0.05.

The confidence interval $(\bar{l} - 0.165, \bar{l} + 0.315)$ does not cover the entire identification region $[E(L), E(L) + 0.15]$ with probability 0.95, but it covers any possible value for μ with such a confidence.

Let us now suppose that we would just had the interval information $X^* \in [L, L + 0.15]$ and we would not know that $X^* - L$ is a constant. The Aumann expectation of the random set $[L, L + 0.15]$ coincides with the above identification region. But, in this case, we would not be able to say that $(\bar{l} - 0.165, \bar{l} + 0.315) = (\bar{l} - 0.165, (\bar{l} + 0.15) + 0.165)$ is a 95% confidence interval and that φ is a 0.05-size test (we would be able to assure that it is a 90% confidence interval).

Another relevant difference with respect to our approach relies on the nature of the confidence regions in both cases. On the one hand, the confidence intervals considered in partial identification problems are closely related to our outer approximations: both kinds of intervals cover the unknown parameter with a prefixed confidence level. On the other hand, our inner approximations do not have any counterpart within the partial identification approach. The role of the inner approximation is crucial under our approach, because it allows us to construct set-valued test functions, as the one considered in Eq. (6), instead of point-valued ones. From a practical point of view, they allow us to distinguish between "no rejection" and "no conclusion", as we illustrate in the following example.

Example 4.5. Consider again the second situation described in Example 4.4: we take a 100 size sample from an observable random variable L with known standard deviation, $\sigma = 1$. We want to test the null hypothesis:

$$H_0 : \mu = \mu_0 \quad \text{against } H_1 : \mu \neq \mu_0,$$

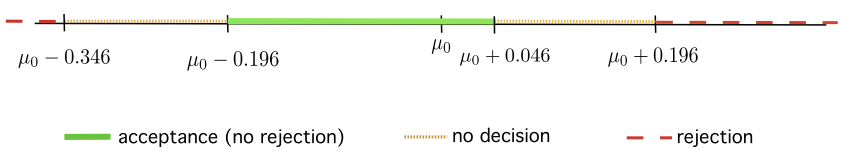
where μ denotes the expectation of a random variable X^* that is known to take values in the interval $[L, L + 0.15]$ (we do not directly observe X^*, but for each observation l_i of L, we can say that the corresponding value x_i^* belongs to the interval $[l_i, l_i + 0.15]$). Following the procedure described in Eq. (6), we can build a pair of outer and inner approximations for the 0.95 confidence interval $(\overline{x^*} - 0.196, \overline{x^*} + 0.196)$. For a specific set-valued realization $\gamma = [l_1, l_1 + 0.15] \times [l_{100}, l_{100} + 0.15]$, they are calculated as follows:

$$\overline{CR}(\gamma) = \cup_{\vec{x} \in \gamma}(\bar{x} - 0.196, \bar{x} + 0.196) = (\bar{l} - 0.196, \bar{l} + 0.346),$$
$$\underline{CR}(\gamma) = \cap_{\vec{x} \in \gamma}(\bar{x} - 0.196, \bar{x} + 0.196) = (\bar{l} - 0.046, \bar{l} + 0.196).$$

Furthermore, we can derive from them the following set-valued test function:

$$D_{\varphi''}(\gamma) = \begin{cases} \{1\} & \text{if } \ \overline{CR}(\gamma) \not\ni \mu_0 \\ \{0\} & \text{if } \ \underline{CR}(\gamma) \ni \mu_0 \\ \{0,1\} & \text{otherwise} \end{cases} = \begin{cases} \{1\} & \text{if } \ \bar{l} > \mu_0 + 0.196 \quad \text{or} \quad \bar{l} < \mu_0 - 0.346 \\ \{0\} & \text{if } \ \mu - 0.196 < \bar{l} < \mu_0 + 0.046 \\ \{0,1\} & \text{otherwise} \end{cases}$$

The use of the inner region allows us to distinguish between the no rejection cases (for which the output of the test function is {0}) from the not concluding situations (where the output of the test function is {0,1}, due to the imprecision of our observations). The distinction in practice is important: they allow us to differentiate the situations where the sample data assure us that the sample belongs to the acceptance region from those samples that, due to the imprecision of our observations, can not be classified in the rejection region, nor in the acceptance region.



More specifically, those samples where the observable sample mean, $\bar{l}$, falls within the interval $(\mu_0 - 0.196, \mu_0 + 0.046)$ belong to the acceptance region with complete certainty. On the contrary, any of the samples satisfying the restriction $\bar{l} \in (\mu_0 - 0.346, \mu_0 - 0.196) \cup (\mu_0 - 0.046, \mu_0 + 0.196)$ is known to belong to the rejection or to the acceptance region. In other words, with this kind of test, we distinguish those samples that "support" the null hypothesis from those samples that, due to the imprecision of our observations, do not allow us to take a decision.

As illustrated at the end of the last example, the more precision in the data, the lower the proportion of samples where the extended α-test is inconclusive. In the next section, we will show that marking and recapture techniques give us additional information and reduce the proportion of samples where the test prevents us from taking a decision.

5. Mark and recapture techniques

In this section, we will consider the set-valued test for imprecise data considered in the previous section. We want to show that this method can be improved if we use mark and recapture techniques when we take a particular sample. More specifically, we will show that marking the sample elements induces a lower proportion of cases where the test is inconclusive. Let us introduce these ideas with some illustrative examples:

Example 5.1. Let us recall Example 4.2. We took one hundred objects from a large population, and weighed them in a coarse scale, obtaining the following frequency table:

γ_i	[5.5, 6.5)	[6.5, 7.5)	[7.5, 8.5)
n_i	27	49	24

Let us compute the bounds of $\widehat{S}_{X^*}$. These can be obtained by solving the box-constrained optimization problem:

$$\min(\max) \frac{1}{99} \sum_{i=1}^{100} \left(x_i - \frac{\sum_{i=1}^{100} x_i}{100} \right)^2, \tag{8}$$

subject to

$$\begin{cases} 5.5 \leqslant x_i < 6.5 & i = 1 \ldots 27, \\ 6.5 \leqslant x_i < 7.5 & i = 28 \ldots 76, \\ 7.5 \leqslant x_i < 8.5 & i = 77 \ldots 100. \end{cases} \tag{9}$$

The minimum and maximum values of the sample variance, given by Eq. (9), are 0.130 and 1.296. However, in Example 4.2, the initial data set (with the precise measurements) showed that there were only four different classes of objects. Now, suppose that this extra information is available, in addition to the information provided by Table 1. That is to say, we know that all objects where $\gamma_i = [5.5, 6.5)$ weight the same (unknown) amount. The same can be said about those objects where $\gamma_i = [7.5, 8.5)$.

This extra information can be accounted by two new equality constraints in the optimization problem:

$$x_1 = x_2 = \cdots = x_{27}, \tag{10}$$

$$x_{77} = x_{78} = \cdots = x_{100}, \tag{11}$$

that reduce the number of variables in the optimization problem from 100 to 51.

Lastly, in case we can mark the members of all the four classes, we can add two more equality constraints:

$$x_{28} = x_{29} = \cdots = x_{51}, \tag{12}$$

$$x_{52} = x_{53} = \cdots = x_{76}, \tag{13}$$

that reduce the optimization problem size to 4 unknowns. In addition, the new bounds of the sample variance are tighter (0.130 and 1.255). More details about this technique, the so called mark-recapture process, will be given at the end of this section.

Example 5.2. Suppose that we have three urns. Each of the urns contains balls which are coloured either red or white. You will select one of the three urns according to the following procedure. You will choose at random a card from a deck of 52 playing

cards. If a queen is chosen, then you will select a ball at random from urn number 1. If you choose an ace, you will select a ball from urn number 2. Otherwise, you will select a ball from urn number 3. You neither know how many balls are in each urn, nor the proportion of red and white balls. But, in fact, each of the urns numbers 1 and 2 only contains one ball. The one in the first urn is white and the one in the second urn is red. Let us now consider two different situations for the third urn:

1. In the first case, there is only one ball, which is red, but it is kept in a box and you cannot open it.
2. In the second case, there are a hundred balls. Half of them are white and half of them are red. Each ball is kept in an individual box that you cannot open.

Consider the random variable $X : \Omega \to \{0,1\}$ defined on the set of balls Ω as follows:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is white,} \\ 0 & \text{if } \omega \text{ is red.} \end{cases}$$

X is a Bernoulli random variable, and the parameter p of success is the probability of selecting a white ball. In the first case $p = 1/13$ and in the second situation, $p = 0.5$, but you do not know it.

Let us consider the following one-sided test about the variance of X:

$$H_0 : \sigma^2 \geqslant 3/16 \quad \text{against } H_1 : \sigma^2 < 3/16.$$

Since $\sigma^2 = p(1-p)$, this test is equivalent to the following one:

$$H_0 : p \in [0.25, 0.75] \quad \text{against } H_1 : p \notin [0.25, 0.75].$$

For all $p \in [0.25, 0.75]$ and all $c \in \mathbb{R}$, we easily check that:

$$\frac{\sqrt{n}(c-0.25)}{\sqrt{0.25 \cdot 0.75}} \geqslant \frac{\sqrt{n}(c-p)}{\sqrt{p \cdot (1-p)}},$$

and

$$\frac{\sqrt{n}(c-0.75)}{\sqrt{0.5 \cdot 0.5}} \leqslant \frac{\sqrt{n}(c-p)}{\sqrt{p \cdot (1-p)}}.$$

Hence, an asymptotic α-test should be determined by the critical region:

$$C = \left\{ (x_1, \ldots, x_n) : \frac{\sqrt{n}(\bar{x}-0.25)}{\sqrt{0.25 \cdot 0.75}} < -z_{1-\alpha/2} \quad \text{or} \quad \frac{\sqrt{n}(\bar{x}-0.75)}{\sqrt{0.5 \cdot 0.5}} > z_{1-\alpha/2}, \right\},$$

where z_p is the q-quantile associated to the standard normal distribution and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ represents the proportion of white balls in the sample. Thus, for instance, if $\alpha = 0.05$ and $n = 100$, then the critical region becomes:

$$C = \{(x_1, \ldots, x_n) : \bar{x} < 0.165 \quad \text{or} \quad \bar{x} > 0.848\}.$$

Let now the multi-valued mapping $\Gamma : \Omega \to \wp(\{0,1\})$ represent your imprecise observation of X. It is defined as follows:

$$\Gamma(\omega) = \begin{cases} \{1\} & \text{if } \omega \text{ is white(and you can see its colour),} \\ \{0\} & \text{if } \omega \text{ is red(and you can see it),} \\ \{0,1\} & \text{if } \omega \text{ is kept in a box.} \end{cases}$$

Suppose that you can take samples of size 100, i.e., you can repeat the above experiment a hundred times (you first select a card and then you select a ball with replacement from the corresponding urn). Then you get an incomplete knowledge about the colour of the 100 balls which is described by the vector $\vec{\gamma} = (\gamma_1, \ldots, \gamma_{100}) \in \wp(\{0,1\})$. You must take a decision about the one-sided test on the basis of this incomplete information. The extension of the above α-test should be represented by $\varphi : \wp(\{0,1\})^n \to \wp(\{0,1\})$ which is defined as follows:

$$\varphi(\vec{\gamma}) = \begin{cases} \{1\} & \text{if } \vec{\gamma} \subseteq C, \\ \{0\} & \text{if } \vec{\gamma} \subseteq C^c \ \{0,1\} \text{if } \vec{\gamma} \cap C \neq \emptyset \quad \text{and} \quad \vec{\gamma} \cap C^c \neq \emptyset. \end{cases}$$

In other words, φ is defined as follows:

$$\varphi(\vec{\gamma}) = \begin{cases} \{1\} & \text{if} \quad \bar{x} \langle 0.165 \quad \text{or} \quad \bar{x} \rangle 0.848, \quad \forall \vec{x} \in \vec{\gamma}, \\ \{0\} & \text{if} \quad \bar{x} \in [0.165, 0.848], \quad \forall \vec{x} \in \vec{\gamma}, \\ \{0,1\} & \text{if} \quad \exists \vec{x}, \vec{x}' \in \vec{\gamma} \quad \text{s.t.} \quad \bar{x} \in [0.165, 0.848]^c \quad \text{and} \quad \overline{x'} \in [0.165, 0.848]. \end{cases}$$

Suppose that, for a specific sequence of 100 selections you have made:

- 8 selections from the first urn.
- 7 selections from the second urn.
- 85 selections from the third urn.

For this particular sample, the class of possible values for the sample mean (the proportion of white balls) is

$$\left\{\bar{x}=\frac{\sum_{i=1}^{100}x_i}{100}:(x_1,\ldots,x_{100})\in\vec{\gamma}\right\}=\left\{\frac{i}{100}:8\leqslant i\leqslant 93\right\}.$$

This set of values satisfies the following conditions:

$$\left\{\frac{i}{100}:8\leqslant i\leqslant 93\right\}\cap[0.165,0.848]\neq\emptyset \quad \text{and}$$
$$\left\{\frac{i}{100}:8\leqslant i\leqslant 93\right\}\cap([0,0.165)\cup(0.848,1])\neq\emptyset.$$

Thus, the extended α-test described by the set-valued function φ should be inconclusive for this particular sample. Nevertheless, mark and recapture techniques should give us enough information to reject the null hypotheses for the first situation described at the beginning of this example. In fact, if we had marked the closed box the first time that we selected it (the first time we chose urn number 3) we should notice that we were selecting the same box the 84 subsequent times. Thus, we should not know the colour of the ball inside, but we should know that it was the same color every time. According to this additional information, there should be only two possible values for the sample proportion $\bar{x}=\frac{\sum_{i=1}^{100}x_i}{100}$. In fact, we should know that it is either $\bar{x}=0.08$ or $\bar{x}=0.93$. Both values belong to the set of values $[0,0.165)\cup(0.848,1]$, so we should reject H_0.

Let us now formalize the above idea of marking individuals to obtain more precise information. Let $\Omega=\{\omega_1,\ldots,\omega_N\}$ be a finite population. Let the random variable $X:\Omega\to\mathbb{R}$ represent some property of the elements of Ω. Let the multi-valued mapping $\Gamma:\Omega\to\wp(\mathbb{R})$ represent an imprecise perception of X. Let $(X_1,\ldots,X_n):\Omega^n\to\mathbb{R}^n$, and $(\Gamma_1,\ldots,\Gamma_n):\Omega^n\to[\wp(\mathbb{R})]^n$ be samples of size n, respectively taken from X and Γ, i.e.

$$X_i(\omega_1,\ldots,\omega_n)=X(\omega_i)\quad\text{and}\quad\Gamma_i(\omega_1,\ldots,\omega_n)=\Gamma(\omega_i),\quad i=1,\ldots,n.$$

$(\Gamma_1,\ldots,\Gamma_n)$ represents imprecise information about the random sample $(X_1,\ldots,X_n)$. But, is it the most informative multi-valued mapping? Or, on the contrary, can we define a more precise multi-valued mapping on Ω^n that describes our information about $(X_1,\ldots,X_n)$?

Let $Y:\Omega\to\mathbb{N}$ be defined as

$$Y(\omega_i)=i,\quad\forall i=1,\ldots,N. \tag{14}$$

Let $(Y_1,\ldots,Y_n):\Omega^n\to\mathbb{R}^n$ represent a sample of size Y taken from Y, i.e., $Y_i(\omega_1,\ldots,\omega_n)=Y(\omega_i)$, $\forall i=1,\ldots,n$. Let us assume that we can observe these sample values because we have previously marked all the elements of the population.

Lemma 5.1. *Let $\vec{\Gamma}':\Omega^n\to\wp(\mathbb{R}^n)$ be the multi-valued mapping defined as follows:*

$$\vec{\Gamma}'(\vec{\omega})=\left\{(x_1,\ldots,x_n):x_i\in\Gamma_i(\vec{\omega}),\quad\forall i=1,\ldots,\quad\text{and}\quad\left[Y_i(\omega)=Y_j(\omega)\Rightarrow x_i=x_j\right]\right\}.$$

Then,

1. $\{(X_1(\vec{\omega}),\ldots,X_n(\vec{\omega}))\}\subseteq\vec{\Gamma}'(\vec{\omega})\subseteq\Gamma_1(\vec{\omega})\times\cdots\times\Gamma_n(\vec{\omega}),\ \forall\vec{\omega}\in\Omega^n$.
2. *In particular, when Γ represents a precise observation of X, i.e., $\Gamma(\omega)=\{X(\omega)\}$ $\forall\omega\in\Omega$, then Γ' represents the same information as $(X_1,\ldots,X_n)$.*

Proof

1.
   - We easily check the inclusion $\{(X_1(\vec{\omega}),\ldots,X_n(\vec{\omega}))\}\subseteq\vec{\Gamma}'(\vec{\omega})$, since, on the one hand, $(X_1(\vec{\omega}),\ldots,X_n(\vec{\omega}))\in\Gamma_1(\vec{\omega})\times\cdots\times\Gamma_n(\vec{\omega})$ and, on the other hand Y is injective and thus $[Y(\omega_i)=Y(\omega_j)\Rightarrow X(\omega_i)=X(\omega_j)]$.
   - The proof of the second inclusion is straightforward.
2. We easily derive from the first part of this lemma that, in such a case $\{(X_1(\vec{\omega}),\ldots,X_n(\vec{\omega}))\}=\vec{\Gamma}'(\vec{\omega})=\Gamma_1(\vec{\omega})\times\ldots\times\Gamma_n(\vec{\omega}),\forall\omega\in\Omega$. □

According to the first point of this lemma, the information accuracy should benefit from marking and recapturing techniques. According to the second point, Γ' represents the initial sample when the perception of X is precise. Let now $\varphi:\mathbb{R}^n\to\{0,1\}$ denote a non-randomized α-test and let us define $T_{\vec{X}}:\Omega^n\to\{0,1\}$, $T_{\vec{\Gamma}}:\Omega^n\to\wp(\{0,1\})$ and $T_{\vec{\Gamma}'}:\Omega^n\to\wp(\{0,1\})$ as follows:

- $T_{\vec{X}}:\Omega^n\to\{0,1\}$ denotes the composition $T_{\vec{X}}=\varphi\circ(X_1,\ldots,X_n)$, i.e.,
  $$T_{\vec{X}}(\vec{\omega})=\varphi(X_1(\vec{\omega}),\ldots,X_n(\vec{\omega}))=\varphi(X(\omega_1),\ldots,X(\omega_n)),\quad\forall\vec{\omega}=(\omega_1,\ldots,\omega_n)\in\Omega^n.$$

- $T_{\vec{\Gamma}} : \Omega^n \to \wp(\{0,1\})$ denotes the multi-valued mapping:

$$T_{\vec{\Gamma}}(\vec{\omega}) = \{\varphi(x_1,\ldots,x_n) : x_i \in \Gamma_i(\vec{\omega}) = \Gamma(\omega_i), \quad \forall i = 1,\ldots,n\}.$$

- $T_{\vec{\Gamma}'} : \Omega^n \to \wp(\{0,1\})$ denotes the multi-valued mapping:

$$T_{\vec{\Gamma}'}(\omega_1,\ldots,\omega_n) = \{\varphi(x_1,\ldots,x_n) : (x_1,\ldots,x_n) \in \Gamma'(\omega_1,\ldots,\omega_n)\}.$$

Corollary 5.1. *According to the above notation, the following inclusions hold:*

- $T_{\vec{X}}(\vec{\omega}) \in T_{\vec{\Gamma}'}(\vec{\omega}) \subseteq T_{\vec{\Gamma}}(\vec{\omega}),\ \forall \vec{\omega} \in \Omega^n$.
- *If* $\Gamma = \{X\}$ *then* $T_{\vec{X}}(\vec{\omega}) = T_{\vec{\Gamma}'}(\vec{\omega}) = T_{\vec{\Gamma}}(\vec{\omega}),\ \forall (\vec{\omega}) \in \Omega^n$.

This means that the extended test associated to $\vec{\Gamma}'$ is as least as precise as the test associated to $\vec{\Gamma}$. They can only report different results when the test associated to $\vec{\Gamma}$ is inconclusive.

Needs:

- The sets of intervals `[L,U]` with `L[i]`$=L(\omega_i)$, `U[i]`$=U(\omega_i)$
- The sample size n
- The value σ_0

Produces:

- The interval of values `T`

Local functions:

`fm((`$a_1,\ldots,a_k$`),(NS[1],...,NS[k]),`σ_0`)` :

return $\dfrac{1}{\sigma_0^2}\left(\sum_{s=1}^{k} a_s^2 \mathtt{NS}[s] - \dfrac{(\sum_{s=1}^{k} a_s \mathtt{NS}[s])^2}{\sum_{s=1}^{k} \mathtt{NS}[s]}\right)$

begin

```
// V is an array of integers of size n
// MARK is an array of integers with the same size as the population
k ← 0; MARK ← 0
for sample ← 1 to n
   i ← index of a random element in the population
   if (MARK[i] = 0) then
        k ← k+1; MARK[i] ← k
   end if
   V[sample] ← i;
end for

// NS is an array of integers of size k
NS ← 0
for i ← 1 to n
   NS[MARK[V[i]] ← NS[MARK[V[i]]+1
end for

// A is matrix of floats of size k × 2
for i ← 1 to n
   A[MARK[V[i]]][1] ← L[V[i]]; A[MARK[V[i]]][2] ← U[V[i]]
end for

// T is found with a suitable numerical optimization algorithm
T ← [ min fm(X,NS,μ0), max fm(X,NS,μ0) ]
     where A[i][1]≤X[i]≤A[i][2], 1≤i≤k
```

end

Fig. 1. Computer algorithm for computing a test for the variance based on set-valued data. If the set T is not contained in the regions stated in Section 5.1, the test is inconclusive; otherwise, a decision can be made.

Remark 5.1. Throughout this section we have assumed that we can mark each individual in the population and most of the times this is not possible in practice: the population may be so large and marking individuals could be very expensive or impractical. Even so, using this technique may be feasible, since we only need to mark the sample elements. Thus, we should observe the values of the ordering variable Y defined in Eq. (14) (more strictly speaking, we observe the values of a bijection of Y) on the sample and we could apply also this technique when the population is not finite. According to these considerations, we extend the classical test of the mean in the following subsection.

5.1. Particular case: variance test for imprecise data

Let us apply these mark-recapture techniques to extend the classical test of the variance for normal populations. Let us take a random sample of size n and let l_i and u_i be the lower and the upper bounds of the interval data. The procedure is as follows:

1. In the first place, we remove all the elements of the sample that appear more than once, leaving one single copy of each. This process can also be thought of as iterating over the sample, then
   (a) mark each element that has not been seen before, and
   (b) update a counter each time a marked element is recaptured (i.e., whenever an already seen element appears again in the sample).

Let us name the k elements that are left $(\omega_{v_1},\ldots,\omega_{v_k})$. The counters of these elements are $n_1,\ldots,n_k$, where n_s is the absolute frequency of ω_{v_s} in the original sample.

2. Next, the range of values of the sample variance is computed. Each possible value of the sample variance can be written as follows:

$$S^2 = \frac{n}{n-1}\left[\frac{\sum_{s=1}^k a_s^2 n_s}{n} - \left(\frac{\sum_{s=1}^k a_s^2 n_s}{n}\right)^2\right] : a_s \in [l_{v_s}, u_{v_s}].$$

Thus, all we know about the sample variance is that it belongs to the class of values:

$$M = \{f(a_1,\ldots,a_k) : (a_1,\ldots,a_k) \in A\}, \quad \text{where}$$

$$f(a_1,\ldots,a_k) = \frac{n}{n-1}\left[\frac{\sum_{s=1}^k a_s^2 n_s}{n} - \left(\frac{\sum_{s=1}^k a_s n_s}{n}\right)^2\right], \quad \text{and}$$

$$A = \{(a_1,\ldots,a_k) \in \mathbb{R} : a_s \in [l_{v_s}, u_{v_s}]\}.$$

In Fig. 1 there is a computer algorithm for performing the calculations explained previously. It is emphasized that the computation of the bounds of M is a numerical problem that depends on finding the extrema of a function. If mark-recapture techniques are used, this function depends on k variables, or else it depends on n variables. Since $k \leqslant n$, mark-recapture techniques require less computer time.

Let us now extend the classical variance test for interval data. Let μ denote the expectation of X. Let T_σ denote the chi-square statistic $T_\sigma = \frac{(n-1)S^2}{\sigma^2}$, where S^2 represents the sample variance, based on a random sample $(X_1,\ldots,X_n)$. Under the assumption of normality, T_σ is a chi-square random variable. In the following table we summarize classical results about this parametric test:

	H_0	H_1	Reject H_0 at level α if
I	$\sigma \leqslant \sigma_0$	$\sigma > \sigma_0$	$T_{\sigma_0}(\omega_1,\ldots,\omega_n) \geqslant \chi_{n-1,1-\alpha}$
II	$\sigma \geqslant \sigma_0$	$\sigma < \sigma_0$	$T_{\sigma_0}(\omega_1,\ldots,\omega_n) \leqslant \chi_{n-1,\alpha}$
III	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$T_{\sigma_0}(\omega_1,\ldots,\omega_n) \leqslant \chi_{n-1,\alpha}$ or $T_{\sigma_0}(\omega_1,\ldots,\omega_n) \geqslant \chi_{n-1,1-\alpha/2}$

If we get interval-valued observations $[l_1,u_1],\ldots,[l_n,u_n]$, the class of possible values of T_{σ_0} is

$$N = \left\{\frac{(n-1)s^2}{\sigma_0} : s^2 \in M\right\}.$$

Thus, for instance, according to the second one-tailed test (test II) we propose the following decision scheme:

- Reject H_0 if $N \subseteq [0, \chi_{n-1,\alpha})$.
- Accept H_0 if $N \subseteq [\chi_{n-1,\alpha}, \infty]$.
- Otherwise, we will say that the test is inconclusive, due to imprecision in measurements.

Example 5.3. Suppose that we take a random sample, with replacement and without marking. The observations of the characteristic under study are not precise, and the elements of the sample are intervals. The size of the sample will be higher than that of the population. For instance, we are performing a bootstrap analysis of the data.

Let us use small numbers, so the calculations can be done by hand. Imagine that we have taken a sample of size $n = 6$ from a population of size 3. This sample is

$$[l_i, u_i]: \quad [0,10] \quad [10,20] \quad [20,30] \quad [0,10] \quad [10,20] \quad [20,30],$$

and we want to test:

$$H_0: \sigma \leqslant 28.6 \quad \text{against} \quad H_1: \sigma > 28.6.$$

In the first place, let us compute the extrema of the statistic $T_{\sigma_0} = \frac{(n-1)S^2}{\sigma_0^2}$, where $\sigma_0 = 28.6$. The lowest bound of T_{σ_0} is $100/\sigma_0^2$, which would be attained if the actual values of the elements of the sample were:

$$10 \quad 15 \quad 20 \quad 10 \quad 15 \quad 20.$$

Conversely, the upper bound of T_{σ_0} is $950/\sigma_0^2$, produced by the values:

$$0 \quad 10 \quad 30 \quad 0 \quad 20 \quad 30.$$

Hence, the interval N is $[0.12, 1.16]$. The test is inconclusive at significance level $\alpha = 0.05$.

Secondly, we are going to apply mark-recapture techniques. Observe that, if one element of the population appears more than once, all their instances are assigned the same interval. By using this extra information, we are able to make stronger assertions about the population. The lower bound of T_{σ_0} will be the same as before, but the upper bound is no longer $950/\sigma_0$, but $933.33/\sigma_0$, for a sample:

$$0 \quad 10 \quad 30,$$

where each element appears twice. Therefore, $N = [0.12, 1.14]$, which is narrower than before. In this case, we have evidence against the null hypothesis.

6. A basic bootstrap test for imprecise data

Bootstrap-based confidence intervals and tests can take advantage of mark-recapture techniques, simplifying the numerical computations and, in certain circumstances, obtaining more conclusive tests from imprecise data.

In this section we will show, with the help of an example, how to extend the basic bootstrap confidence limits for a parameter [20] to the case of interval data, applying mark-recapture techniques. A test will be derived from this confidence limits, with the help of Eq. (6).

We will summarize first the concept of basic bootstrap confidence region. Let $X = (X_1, \ldots, X_n)$ be a random sample, and let $T = g(X)$ be an estimator of a scalar parameter θ. Let $\vec{x} = (x_1, \ldots, x_n)$ be a realization of the sample X, and let $t = g(\vec{x})$ be the punctual estimation of θ for the sample realization $\vec{x}$. Let $\widehat{F}$ be the empirical distribution function:

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} H(x - x_i), \tag{15}$$

where $H(u)$ is the unit step function that jumps from 0 to 1 at $u = 0$. Lastly, let $T^* = g(X^*)$ be the same statistic based on a sample $X^* = (X_1^*, \ldots, X_n^*)$, where the X_i^* are independently sampled from the distribution $\widehat{F}$.

The basic intervals are based on the idea that the distribution of $T - \theta$ mimics that of $T^* - T$. Let $k_{\alpha/2}(\vec{x})$ and $k_{1-\alpha/2}(\vec{x})$ be the $\alpha/2$ and $1 - \alpha/2$ quantiles of the distribution of T^*, respectively. The basic bootstrap interval expression, for a particular sample $\vec{x} = (x_1, \ldots, x_n)$ is

$$\left(2g(\vec{x}) - k_{1-\alpha/2}(\vec{x}), 2g(\vec{x}) - k_{\alpha/2}(\vec{x})\right), \tag{16}$$

$k_{\alpha/2}(\vec{x})$ and $k_{1-\alpha/2}(\vec{x})$ are estimated by computer simulation: we generate R realizations $\vec{x}_r^*$ of X^*, resampling $\vec{x}$ with replacement, and estimate the cumulative probability $G(u) = P(T \leqslant u)$ with

$$\widehat{G}_R(u) = \frac{1}{R}\sum_{r=1}^{R} H(u - t_r^*). \tag{17}$$

If R is high enough, we approximate the quantile h, $k_h(\vec{x})$, by

$$k_h^*(\vec{x}) = \max\{u | \widehat{G}_R(u) \leqslant h\} \tag{18}$$

which is the $h(R+1)$ th ordered value of t^*. R is chosen so that $h(R+1)$ is an integer. A realization of the approximated basic bootstrap confidence region is, therefore, the interval:

$$CR(\vec{x}) = \left(2g(\vec{x}) - k_{1-\alpha/2}^*(\vec{x}), 2g(\vec{x}) - k_{\alpha/2}^*(\vec{x})\right). \tag{19}$$

We will construct a parametric test for a simple null hypothesis upon this confidence region. As mentioned in Section 4.2.1, when CR is a $1 - \alpha$-confidence interval, the mapping $\varphi_\alpha'': \mathbb{R}^n \to \{0, 1\}$:

$$\varphi''_\alpha(\vec{x}) = \begin{cases} 0 & \text{if} \quad \theta_0 \in CR(\vec{x}), \\ 1 & \text{otherwise}, \end{cases}$$

is an α-test of $H_0 : \theta(X^*) = \theta_0$ against $H_1 : \theta(X^*) \neq \theta_0$. For making a decision on the basis of a sample of n imprecise measurements such that $\vec{x} \in \gamma$, we will use the procedure described in Eq. (6):

$$D''_\varphi(\gamma) = \begin{cases} \{1\} & \text{if } \theta_0 \notin \overline{CR}(\gamma) \\ \{0\} & \text{if } \theta_0 \in \underline{CR}(\gamma) \\ \{0,1\} & \text{otherwise}. \end{cases} \tag{20}$$

In this case, $\underline{CR}(\gamma)$ and $\overline{CR}(\gamma)$ can be expressed as follows:

$$\underline{CR}(\gamma) = \cap_{\vec{x}\in\gamma}\Big(2g(\vec{x}) - k^*_{1-\alpha/2}(\vec{x}), 2g(\vec{x}) - k^*_{\alpha/2}(\vec{x})\Big), \tag{21}$$

$$\overline{CR}(\gamma) = \cup_{\vec{x}\in\gamma}\Big(2g(\vec{x}) - k^*_{1-\alpha/2}(\vec{x}), 2g(\vec{x}) - k^*_{\alpha/2}(\vec{x})\Big). \tag{22}$$

Obtaining the limits of the above intervals is a hard computational problem, because there is not a general procedure for obtaining the solution of a box-constrained optimization of a non-differentiable function. We could use metaheuristics, nonlinear programming techniques, or Monte-Carlo simulation, neither of which has an admissible cost if the sample size is large.

However, computing the approximation:

$$\underline{\underline{CR}}(\gamma) = \left[2 \cdot \max_{\vec{x}\in\gamma} g(\vec{x}) - \min_{\vec{x}\in\gamma} k^*_{1-\alpha/2}(\vec{x}), 2 \cdot \min_{\vec{x}\in\gamma} g(\vec{x}) - \max_{\vec{x}\in\gamma} k^*_{\alpha/2}(\vec{x})\right], \tag{23}$$

$$\overline{\overline{CR}}(\gamma) = \left[2 \cdot \min_{\vec{x}\in\gamma} g(\vec{x}) - \max_{\vec{x}\in\gamma} k^*_{1-\alpha/2}(\vec{x}), 2 \cdot \max_{\vec{x}\in\gamma} g(\vec{x}) - \min_{\vec{x}\in\gamma} k^*_{\alpha/2}(\vec{x})\right], \tag{24}$$

is a feasible problem, and we can take advantage of mark-recapture techniques for computing $T(\vec{x})$ and k^*. In the following example we illustrate the construction of an approximated bootstrap test for the same data that we have used in Example 5.3, and also show the differences between the solution of Eqs. (21) and (22) and Eqs. (23) and (24), this last ones with and without mark-recapture techniques.

Example 6.1. Let us consider the sample:

$$[l_i, u_i]: \quad [0,10] \quad [10,20] \quad [20,30] \quad [0,10] \quad [10,20] \quad [20,30].$$

We will provide inner and outer approximations for the basic bootstrap confidence interval based in the standard deviation. We just aim to exemplify how mark-recapture techniques allow us to find more accurate approximations. We approximate the so-called "basic bootstrap confidence interval" construction [20] just for simplicity, and we choose the sample variance just for the sake of unity w.r.t. the examples provided in Section 5. The analysis of the properties of this specific random interval fall beyond the scope of the paper.

Let S be the sample standard deviation, and let us take $R = 99$ resamples with replacement, $\gamma^*_1, \ldots, \gamma^*_R$, of the imprecise sample γ. We need to compute the tightest bounds for the quantiles $k^*_{0.025}$ and $k^*_{0.975}$; first, we compute the bounds for the cumulative empirical distribution function:

$$\overline{G_{99}}(u) = \frac{1}{99}\sum_{r=1}^{99} H\left(u - \min_{\vec{x}^*\in\gamma^*_r}\{S(\vec{x}^*)\}\right), \tag{25}$$

$$\underline{G_{99}}(u) = \frac{1}{99}\sum_{r=1}^{99} H\left(u - \max_{\vec{x}^*\in\gamma^*_r}\{S(\vec{x}^*)\}\right). \tag{26}$$

Observe that the computation of $\min_{\vec{x}^*\in\gamma^*_r}\{S(\vec{x}^*)\}$ and $\max_{\vec{x}^*\in\gamma^*_r}\{S(\vec{x}^*)\}$ involves an optimization problem similar to that mentioned in the preceding section. In this case we can take advantage of the mark-recapture techniques because we know that the duplicate elements originated in the resampling with replacement are imprecise observations of the same value.

In Fig. 2 we have plotted the bounds of these cumulative empirical distribution function. The bounds of the quantile k^*_h are determined by the intersection between $\underline{G_{99}}$, $\overline{G_{99}}$ and the horizontal line at height h. Without mark-recapture:

$$k^*_{0.025} \in [0, 5.47], \quad k^*_{0.975} \in [5.17, 15.49],$$

and therefore:

$$\underline{\underline{CR}} = \emptyset, \quad \overline{\overline{CR}} = [0, 27.57].$$

Using mark-recapture:

$$k^*_{0.025} = 0, \quad k^*_{0.975} \in [5.17, 15.49],$$

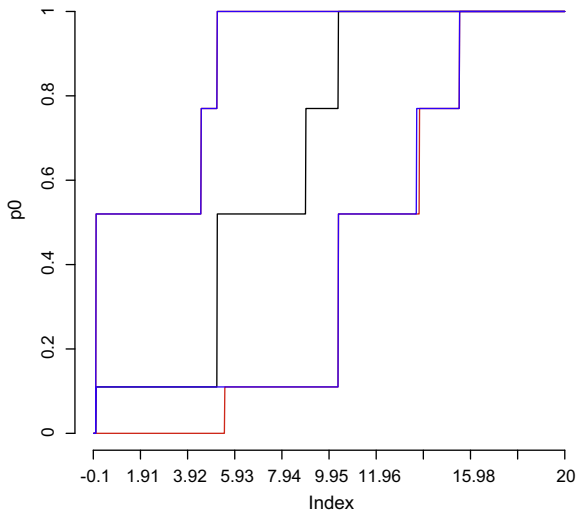


Fig. 2. Bounds of the cumulative empirical distribution function (cedf) G_{99} in Example 6.1. Black: bootstrap cedf originated in the midpoints of the data. Blue: $\overline{\overline{G_{99}}}$ and $\underline{\underline{G_{99}}}$, computed with mark-recapture techniques. Red: $\underline{\underline{G_{99}}}$, computed without mark-recapture. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and therefore:

$$\underline{\underline{CR}} = \emptyset, \quad \overline{\overline{CR}} = [0, 27.32].$$

Observe that, in the first case, the test is inconclusive for values of θ_0 between 27.32 and 27.57, where the second test rejects the null hypothesis. The pseudocode of the algorithm is included in Fig. 3.

We have also computed a Monte-Carlo approximation to Eqs. (21) and (22), thus we can judge the quality of the approximation in Eqs. (23) and (24). Even for this oversimplified problem, the bounds do not become stable until 50000 samples are drawn. Lastly, it is remarked that the final value is not too different from the approximate solution shown before.

Needs:

- The sets of intervals `[XL,XU]`
- The sample size n
- The value θ_0

Produces:

- intervals $\underline{\underline{\texttt{CR}}}$ and $\overline{\overline{\texttt{CR}}}$

begin

`SL,SU` ← bounds of the standard deviation of `[XL,XU]`
`for R` ← `1 to R`
  γ_r ← resample with replacement of `[XL,XU]`
  `SL*[r],SU*[r]` ← bounds of the standard deviation of γ_r
`end for`
`GL(u)` ← function $\frac{1}{R}\sum_{r=1}^{R} I\{\texttt{SL*[r]} \leq u\}$
`GU(u)` ← function $\frac{1}{R}\sum_{r=1}^{R} I\{\texttt{SU*[r]} \leq u\}$
$\texttt{KL}_{\alpha/2}$ ← find zero of $(\texttt{GU} - \alpha/2)$
$\texttt{KU}_{\alpha/2}$ ← find zero of $(\texttt{GL} - \alpha/2)$
$\texttt{KL}_{1-\alpha/2}$ ← find zero of $(\texttt{GU} - 1 + \alpha/2)$
$\texttt{KL}_{1-\alpha/2}$ ← find zero of $(\texttt{GL} - 1 + \alpha/2)$
$\underline{\underline{\texttt{CR}}}$ ← $[2\cdot\texttt{SU} - \texttt{KL}_{1-\alpha/2},\ 2\cdot\texttt{SL} - \texttt{KU}_{\alpha/2}]$
$\overline{\overline{\texttt{CR}}}$ ← $[2\cdot\texttt{SL} - \texttt{KU}_{1-\alpha/2},\ 2\cdot\texttt{SU} - \texttt{KL}_{\alpha/2}]$

end

Fig. 3. Computer algorithm for computing a bootstrap test for the variance based on set-valued data.

Monte-Carlo samples	1000	10000	50000
$\overline{CR}$	[0,26.56]	[0,26.85]	[0,27.22]

7. Concluding remarks and open problems

We have shown how set-valued tests for set-valued data can benefit from mark-recapture techniques. These techniques seem to be especially useful when the population is small enough that recaptures are common. In such a case, sample information about a particular statistic is much more precise when we mark individuals as we select them from the population. This is the point that we have tried to illustrate in Example 5.2. In the first situation described in the example, the population size is 3 and the sample size is 100. If we do not mark the balls, the lower and upper bounds for the sample variance are 0.065 and 0.253, respectively. But, when we mark them, the upper bound is reduced to 0.074. In Examples 5.3 and 6.1 the differences are less remarkable, since the sample size is $n = 6$, instead of $n = 100$. Example 5.2 illustrates a very special situation. In some applications it is not possible to mark individuals as they are being selected from the population (it can be very expensive, or maybe we have not designed the sampling procedure). Furthermore, extensions of classical tests do not apply in some problems with imprecise data, because we cannot verify the hypotheses that are needed to do so. Nonetheless, we can take advantage from marking procedures in such cases. Even when the sample elements are not marked as they are taken from the population, we can mark them afterwards. In this way, we preserve some relevant information. In particular, this procedure can be useful when applying resampling techniques to determine (estimate) the distribution of statistics, as we have illustrated in Section 6. If we deal with imprecise data, we will preserve more information if we mark the sample individuals before resampling. Moreover, the use of marking techniques is justified, even when the information gain is not noticeable, since the computation time is reduced in most cases. When dealing with interval data, taking a decision involves the calculation of a minimum and a maximum of a function defined on a class of feasible solutions. Using mark-recapture techniques reduces this class, so the algorithm tends to end earlier.

The particular case considered in Section 5.1 concerns a parametric test. Nevertheless, applying mark-recapture techniques for finding confidence regions seems to be straightforward. Furthermore, we could apply them in non-parametric inference problems. In particular, we think that mark-recapture techniques could be especially useful in testing stochastic independence and linear correlation. Let us first note that a high (linear or not) dependence between two attributes (X^*, Y^*) does not imply (neither is implied by) a high dependence between the random intervals, $[L_X, U_X]$, $[L_Y, U_Y]$, regarded as imprecise observations of the attributes. For instance, the total ignorance about the values of an attribute would be represented by means of a constant random interval (focussed on the range of possible values) which is stochastically independent from any other random interval, even when the true values of the attribute have some variability. Conversely, when, for instance, the amplitudes of both random intervals are not constant (i.e., the degree of precision of our observations can vary with time) but both amplitudes are somehow related, two independent attributes could be represented by means of a pair of stochastically dependent random intervals. More detailed discussions are given in [3,9,15,16,24]. So, testing independence between $[L_X, U_X]$ and $[L_Y, U_Y]$ (which basically consists in testing stochastic independence between the random vectors (L_X, U_X) and (L_Y, U_Y)) does not necessarily give any insight about the dependence relations between X^* and Y^*. To detect them, we should actually start from a classical independence test (like a chi-square test, for instance). Then we should consider the extended set-valued test for imprecise data. We suspect that imprecision in measurements has a strong influence on this kind of tests. Thus, independence tests should be frequently inconclusive. Hence, we think that mark-recapture techniques prevent us from propagating more and more imprecision, and allow us to draw more conclusive decisions.

In the near future, we plan to extend the ideas given in this paper to the case where the imprecise observations are described by means of fuzzy sets, and also to apply some of these concepts in practical problems [28,49–52]. These studies will be related to the second-order model introduced in [14,17] for fuzzy random variables. This model is based on the possibilistic interpretation of fuzzy sets [13]. But let us notice that there exist in the literature two additional interpretations of fuzzy random variables, different from the representation of the imprecise observation a standard random variable. On the one hand, Puri and Ralescu claim in [46] that the observations of some random experiments do not consist of numerical outputs, but are represented by vague linguistic terms. According to this idea, the fuzzy random variable is a measurable function, in the classical sense, between a certain σ-algebra of events in the original space and a σ-algebra defined over a class of fuzzy subsets in the final space. On the other hand, in [3,10,12,41,42] fuzzy random variables (and, in particular, random sets) are viewed as conditional possibility measures. More specifically, in this approach, it is assumed that there is a sequence of two random experiments on Ω and $\mathbb{R}$, respectively. The probability measure, P, ruling the first one is completey determined. On the other hand, the other experiment is only known via a fuzzy relation which assigns, to each outcome ω of the first subexperiment in the sample set Ω, the fuzzy set $\widetilde{X}(\omega)$ of possible outcomes of the second experiment. Its membership function is a possibility distribution $\widetilde{X}(\omega)(\cdot) = \pi(\cdot \mid \omega)$ that models knowledge about the relationship between the outcome ω of the first sub-experiment and the possible outcomes of the second one. $\pi(\cdot|\omega)$ is called a conditional possibility distribution: if the result of the first experiment is ω, then the possibility degree of $x \in B$ occurring in the second one is $\Pi(B \mid \omega) = \sup_{x \in B} \widetilde{X}(\omega)(x)$. The mark-recapture technique here proposed is not compatible with any of those interpretations of random sets and fuzzy random variables.

Acknowledgements

This work has been supported by Spanish MEC–FEDER Grants TIN2010-20900-C04-01 and TIN2008-06681-C06-04. This financial support is gratefully acknowledged.

References

[1] D.W.K. Andrews, G. Soares, Inference for parameters defined by moment inequalities using generalized moment selection, Econometrica 78 (2010) 119–157.
[2] J. Aumann, Integral of set valued functions, Journal of Mathematical Analysis and Applications 12 (1965) 1–12.
[3] C. Baudrit, I. Couso, D. Dubois, Joint propagation of probability and possibility in risk analysis: towards a formal framework, International Journal of Approximate Reasoning 45 (2007) 82–105.
[4] A. Beresteanu, F. Molinari, Asymptotic properties for a class of partially identified models, Econometrica 76 (2007) 763–814.
[5] L. Billard, E. Diday, Symbolic Data Analysis: Definitions and Examples. <http://www.stat.uga.edu/faculty/LYNNE/tr_symbolic.pdf>.
[6] A. Blanco, N. Corral, G. González-Rodríguez, M.A. Lubiano, Some properties of the d_k-variance for interval-valued randoms sets, in: Fourth International Workshop on Soft Methods in Probability and Statistics (SMPS 2008), Toulouse, France, 2008.
[7] A. Castaldo, F. Maccheroni, M. Marinacci, Random correspondences as bundles of random variables, Sankhya 66 (2004) 409–427.
[8] V. Chernohouz, H. Hong, E. Tamer, Estimation and confidence regions for parameter sets in econometric models, Econometrica 75 (2007) 1243–1284.
[9] I. Couso, Independence concepts in evidence theory, in: Fifth International Symposium on Imprecise Probabilities: Theory and Applications (ISIPTA 2007), Prague, Czech Republic, 2007.
[10] I. Couso, D. Dubois, On the variability of the concept of variance for fuzzy random variables, IEEE Transactions on Fuzzy Systems 17 (2009) 1070–1080.
[11] I. Couso, D. Dubois, S. Montes, L. Sánchez, On various definitions of the variance of a fuzzy random variable, in: Fifth International Symposium on Imprecise Probabilities: Theory and Applications (ISIPTA 2007), Prague, Czech Republic, 2007, pp. 135–144.
[12] I. Couso, E. Miranda, G. de Cooman, A possibilistic interpretation of the expectation of a fuzzy random variable, in: M. López-Díaz, M.A. Gil, P. Grzegorzewski, O. Hryniewicz, J. Lawry (Eds.), Soft Methodology and Random Information Systems, Springer, Berlin, 2004, pp. 133–140.
[13] I. Couso, S. Montes, P. Gil, The necessity of the strong alpha-cuts of a fuzzy set, International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 9 (2001) 249–262.
[14] I. Couso, S. Montes, P. Gil, Second order possibility measure induced by a fuzzy random variable, in: C. Bertoluzza, M.A. Gil, D.A. Ralescu (Eds.), Statistical Modeling, Analysis and Management of Fuzzy Data, Springer, Heidelberg, 2002, pp. 127–144.
[15] I. Couso, S. Moral, Independence concepts in evidence theory, International Journal of Approximate Reasoning 51 (2010) 748–758.
[16] I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities, Risk, Decision and Policy 5 (2000) 165–181.
[17] I. Couso, L. Sánchez, Higher order models for fuzzy random variables, Fuzzy Sets and Systems 159 (2008) 237–258.
[18] I. Couso, L. Sánchez, Inner and outer fuzzy approximations of confidence intervals, Fuzzy Sets and Systems, submitted for publication.
[19] I. Couso, L. Sánchez, P. Gil, Imprecise distribution function associated to a random set, Information Sciences 159 (2004) 109–123.
[20] A.C. Davison, D.V. Hinkley, Bootstrap Methods and Their Applications, Cambridge Series in Statistics and Probabilistic Mathematics, Cambridge University Press, 2003.
[21] G. Debreu, Integration of correspondences, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, part 1, 1967, pp. 351–372.
[22] T. Denoeux, M.H. Masson, P.A. Hebert, Nonparametric rank-based statistics and significance test for fuzzy data, Fuzzy Sets and Systems 153 (2005) 1–28.
[23] Y. Feng, L. Hu, H. Shu, The variance and covariance of fuzzy random variables and their applications, Fuzzy Sets and Systems 120 (2001) 487–497.
[24] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, L. Ginzburg, Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty, Technical Report, SAND2007-0939, Unlimited Release, 2007.
[25] P. Filzmoser, R. Viertl, Testing hypotheses with fuzzy data: the fuzzy p-value, Metrika 59 (2004) 21–29.
[26] J.L. Horowitz, C.F. Manski, Imprecise identification for incomplete data, in: Second International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2001), Ithaca, New York, 2001.
[27] G.W. Imbens, C.F. Manski, Confidence intervals for partially identified parameters, Econometrica 72 (2004) 1845–1857.
[28] M.J. del Jesus, F. Hoffmann, L. Junco, L. Sánchez, Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms, IEEE Transactions on Fuzzy Systems 12 (2004) 296–308.
[29] D.G. Kendall, Foundations of a theory of random sets, in: E.F. Harding, D.G. Kendall (Eds.), Stochastic Geometry, Wiley, New York, 1974, pp. 322–376.
[30] R. Körner, Linear Models with Fuzzy Random Variables, Ph.D. Thesis, 1997.
[31] R. Körner, On the variance of fuzzy random variables, Fuzzy Sets and Systems 92 (1997) 83–93.
[32] R. Körner, An asymptotic α-test for the expectation of random fuzzy variables, Journal of Statistical Planning and Inference 83 (2000) 331–346.
[33] R. Kruse, On the variance of random sets, Journal of Mathematical Analysis and Applications 122 (1987) 469–473.
[34] R. Kruse, K.D. Meyer, Statistics with vague data, D. Reidel Publishing Company, Dordrecht, 1987.
[35] C.F. Manski, Partial identification with missing data: concepts and findings, International Journal of Approximate Reasoning 39 (2005) 151–165.
[36] G. Matheron, Random Sets and Integral Geometry, Wiley, New York, 1975.
[37] E. Miranda, I. Couso, P. Gil, Study of the probabilistic information of a random set, in: Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2003), Lugano, Switzerland, 2003.
[38] E. Miranda, I. Couso, P. Gil, Random intervals as a model for imprecise information, Fuzzy Sets and Systems 154 (2005) 386–412.
[39] E. Miranda, I. Couso, P. Gil, Random sets as imprecise random variables, Journal of Mathematical Analysis and Applications 307 (2005) 32–47.
[40] E. Miranda, I. Couso, P. Gil, Approximations of upper and lower probabilities by measurable selections, Information Sciences 180 (2010) 1407–1417.
[41] E. Miranda, G. de Cooman, I. Couso, Imprecise probabilities induced by multi-valued mappings, in: Ninth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002), Annecy, France, 2002.
[42] E. Miranda, G. de Cooman, I. Couso, Lower previsions induced by multi-valued mappings, Journal of Statistical Planning and Inference 133 (2005) 173–197.
[43] M. Montenegro, M.R. Casals, A. Colubi, M.A. Gil, Testing "two sided" hypothesis about the mean of an interval-valued random set, in: Fourth International Workshop on Soft Methods in Probability and Statistics (SMPS 2008), Toulouse, France, 2008.
[44] H.T. Nguyen, On random sets and belief functions, Journal of Mathematical Analysis and Applications 65 (1978) 531–542.
[45] A.M. Palacios, L. Sánchez, I. Couso, Diagnosis of dyslexia with low quality data with genetic fuzzy systems, International Journal of Approximate Reasoning 51 (2010) 993–1009.
[46] M. Puri, D. Ralescu, Fuzzy random variables, Journal of Mathematical Analysis and Applications 114 (1986) 409–422.
[47] V.K. Rohatgi, An Introduction to Probability Theory and Mathematical Statistics, John Wiley and Sons, USA, 1976.
[48] J.P. Romano, A.M. Shaikh, Inference for identifiable parameters in partially identified econometric models, Journal of Statistical Planning and Inference 138 (2008) 2786–2807.
[49] L. Sánchez, I. Couso, Advocating the use of imprecisely observed data in genetic fuzzy systems, IEEE Transactions on Fuzzy Systems 15 (2007) 551–562.

[50] L. Sánchez, I. Couso, J. Casillas, A multiobjective genetic fuzzy system with imprecise probability fitness for vague data, in: Second International Symposium on Evolving Fuzzy Systems 2006 (EFS 2006), Ambleside Lake District, UK, 2006, pp. 131–136.
[51] L. Sánchez, I. Couso, J. Casillas, Genetic learning of fuzzy rules based on low quality data, Fuzzy Sets and Systems 160 (2009) 2524–2552.
[52] L. Sánchez, M.R. Suárez, J.R. Villar, I. Couso, Mutual information-based feature selection and partition design in fuzzy rule-based classifiers from vague data, International Journal In Approximate Reasoning 49 (2008) 607–622.
[53] J. Stoye, Statistical inference for interval identified parameters, in: Sixth International Symposium on Imprecise Probabilities: Theory and Applications (ISIPTA 2009), Durham, UK, 2009.