entropy

MDPI

Review

A Confidence Set Analysis for Observed Samples: A Fuzzy Set Approach

José Alejandro González [1], Luis Mauricio Castro [2], Víctor Hugo Lachos [3] and Alexandre Galvão Patriota [4,*]

1 Departamento de Matemática y Estadística, Universidad de Playa Ancha, Valparaíso 2340000, Chile; jgonzalez@upla.cl
2 Departamento de Estadística and CI^{2}MA, Universidad de Concepción, Concepción 4030000, Chile; luiscastroc@udec.cl
3 Departamento de Estatística, Universidade Estadual de Campinas, Campinas 13083-859, Brazil; hlachos@ime.unicamp.br
4 Departamento de Estatística, IME, Universidade de São Paulo, São Paulo 05508-090, Brazil
* Correspondence: patriota@ime.usp.br; Tel.: +55-11-3091-6188; Fax: +55-11-3091-6130

Academic Editors: Julio Stern, Adriano Polpo and Antonio M. Scarfone
Received: 20 March 2016; Accepted: 25 May 2016; Published: 30 May 2016

Abstract: Confidence sets are generally interpreted in terms of replications of an experiment. However, this interpretation is only valid before observing the sample. After observing the sample, any confidence sets have probability zero or one to contain the parameter value. In this paper, we provide a confidence set analysis for an observed sample based on fuzzy set theory by using the concept of membership functions. We show that the traditional *ad hoc* thresholds (the confidence and significance levels) can be attained from a general membership function. The applicability of the newly proposed theory is demonstrated by using well-known examples from the statistical literature and an application in the context of contingency tables.

Keywords: confidence sets; fuzzy sets; membership function; possibility theory

1. Introduction

Quantities of interest are typically surrounded by a number of uncertain events. According to statistical reasoning, probability measures are employed to model uncertain events and to make inferences over the quantities of interest. The observed relevant information is contained in the observed sample [1]. The statistical model is formally written by the triplet: $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, where $\mathcal{X} \subseteq \mathbb{R}^n$ is the sample space, $\mathcal{F}$ is the associated σ-field and $\mathcal{P} = \{P^{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is a family of sampling probabilities indexed by a parameter $\boldsymbol{\theta}$, where $\Theta \subset \mathbb{R}^k$, with $k < \infty$, is a non-empty set called the parameter space. The quantities of interest are connected with the parameter $\boldsymbol{\theta}$, e.g., the expectation of some random quantity defined in the statistical model.

The inferential process about $\boldsymbol{\theta}$ involves a summary of the information provided by the observed *data* using (minimal) sufficient statistics and their respective induced models that concentrate the statistical relevant information. There are essentially two types of estimation theories, namely, point and set estimation theories; this paper focuses on the latter. For the univariate case, Neyman [2] provided a theory of confidence intervals, which is based on a random interval $\theta_1(X) \leq \theta \leq \theta_2(X)$ such that its probability is greater than (or equal to) to a given predefined value $\gamma = 1 - \alpha < 1$ (confidence level), where X is the random sample. The most frequent interpretation states that if the experiment is repeated and a confidence interval is computed for each experiment, then the parameter θ is expected to lie at least in $100\gamma\%$ of those observed confidence intervals. However, in practice, the experiment is repeated once and just one confidence set is observed. This observed confidence set

$\theta_1(x) \leq \theta \leq \theta_2(x)$ contains non-random values, since x is the observed value of X, so, the probability that this observed confidence set contains any specific point or region will always be zero or one [3]. Therefore, after observing the sample the confidence sets cannot be interpreted in terms of frequencies (as Neyman proposed in 1935 [2]).

Fuzzy set theory developed by Zadeh [4,5] allows generating possibility distributions by using confidence sets (see for example [6–8]). Probability measures are dominated by possibility measures in the following sense: Events with zero possibility must have zero probability, however not all events with positive possibility have positive probability [9]. That is, in some cases, some events with positive possibility do have zero probability. Therefore, possibility measures can provide an information not featured by probability measures. We will show that, for a given observed sample, the related possibility distribution provides information about the structure of confidence sets.

The main contribution of our paper is to show how to generate a fuzzy number from a given confidence set and therefore infer about some parameter $\boldsymbol{\theta}$ under the light of fuzzy set theory. Although this approach has already been discussed in the literature (see [7,8,10]), our approach is more general (e.g., the confidence region is formally defined, the parameter space is multidimensional, *etc.*) and oriented to the statistical community. In this context, the proposal presented in this paper is based on a general membership function proposed in Patriota [11]. As a consequence of this characterization, properties and comparisons of confidence sets are discussed under the scope of fuzzy set theory but from a statistical point of view.

The paper is organized as follows. In Section 2, a review of fuzzy set theory for statisticians is provided. Section 3 focuses on the connection between confidence sets and fuzzy sets through a membership function. Section 4 presents some examples of the results obtained and Section 5 provides an application to a real dataset. Finally, in Section 6, a discussion about different proposals existing in the literature for relating confidence theory and fuzzy theory is presented. Section 7 ends the paper with some remarks and conclusions.

2. Preliminaries

2.1. A Brief Review of Fuzzy Theory

Fuzzy set theory provides mathematical treatment of some vague linguistic terms such as "about", "around", "close", "short", among others. From the fuzzy theory viewpoint, numbers are idealizations of imprecise information expressed by means of numerical values. For example, when the height of an individual is measured, a numerical value is registered including some inaccuracies. Such inaccuracies may have been caused by the measurement instruments, human limitations, rounding, or biased prior information among many other causes. If the "real" value of the height is represented by the number h, maybe it would be more correct to say that the value of the height is approximately and not exactly h [12], the word "approximately" is imprecise and can be modeled by fuzzy theory. As was noted by Coppi *et al.* [13], fuzzy theory can provide an additional value to statistical methods because of the uncertainty inherent to the observable world and its associated information sources are combined beyond the traditional probability theory. For example, Tanaka *et al.* [14] introduced the concept of fuzzy regression while Wünsche *et al.* [15] characterized the least squares method for fuzzy random variables and Arabpour and Tata [16] developed some theoretical elements regarding parameter estimation in fuzzy regression models. The connection between the estimation of parameters and fuzzy theory has been studied by several authors. Geyer and Meeden [17] established a relation between the concept of p-value and fuzzy structures and Parchami *et al.* [18] introduced the concept of fuzzy confidence intervals. On the other hand, Casals *et al.* [19] studied fuzzy decision problems by relating the concepts of hypothesis testing and fuzzy information nature. Saade and Schwarzlander [20] and Saade [21] proposed a characterization of fuzzy hypothesis testing while Watanabe and Imaizumi [22] related the concepts of hypothesis test statistics and fuzzy hypotheses. Arnold [23,24] related the concept of

fuzzy hypothesis testing with conventional methods of real *data* analysis. Taheri and Behboodian [25] generalized the Neyman–Pearson approach for hypothesis testing under the fuzzy point of view and Filzmoser and Viertl [26] introduced the concept of fuzzy p-value for statistical hypotheses using fuzzy *data*. Recently, Patriota [11] provided an evidence measure for testing null hypotheses that is intrinsically related to fuzzy theory.

2.2. Overview of Fuzzy Theory

In order to make our paper self-contained, we provide an overview of fuzzy theory in this section. We only use some concepts and terminology of fuzzy set theory, mainly based on the works of [5,27,28].

Definition 1. Let $\Omega \subseteq \mathbb{R}^k$ be a non-empty subset of the k-dimensional Euclidean space. A fuzzy set $\widetilde{A}$ is a set of ordered pairs $\widetilde{A} = \left\{ (\boldsymbol{\omega}, \mu_{\widetilde{A}}(\boldsymbol{\omega})) : \boldsymbol{\omega} \in \Omega \right\}$, where $\mu_{\widetilde{A}} : \Omega \to [0,1]$ is called the membership function for the fuzzy set $\widetilde{A}$. In addition, the empty fuzzy set $\widetilde{\varnothing}$ is characterized by $\mu_{\widetilde{\varnothing}}(\boldsymbol{\omega}) = 0$ for all $\boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^k$.

The fuzzy set theory extends the traditional set theory by relaxing the concept of membership of elements in their respective sets. On the one hand, in the ordinary set theory it is considered that $\boldsymbol{\omega} \in A$ (membership one) or $\boldsymbol{\omega} \notin A$ (membership zero), that is, it is a binary operation. On the other hand, fuzzy set theory considers a degree of membership that ranges over the interval $[0,1]$, that is, $\boldsymbol{\omega}$ is a member of A with a certain degree and this same element $\boldsymbol{\omega}$ is a member also of A^c with another degree. Probability theory is built on the usual set theory and provides a number in $[0,1]$ to describe the degree of uncertainty that $\boldsymbol{\omega} \in A$.

The main difference between probability theory and fuzzy theory lies in the definition of a set: the former considers traditional sets and the latter considers fuzzy sets. As will be seen in this section, the properties of fuzzy sets are very different from those of traditional sets. The applicability of fuzzy sets is enormous in language modeling [29], image analysis [30] among many others. One simple example is that an object with gray color has a degree of blackness and a degree of whiteness, so it would be much more informative to model this phenomenon inside the fuzzy set framework setting membership degrees than by setting a binary membership (traditional set theory).

From Definition 1, we can represent an ordinary set by using fuzzy notation. For instance, if $\Omega = \mathbb{R}^k$, then any usual subset $B \subseteq \mathbb{R}^k$ is represented by setting $\mu_{\widetilde{B}}(\boldsymbol{\omega}) = 1$ for all $\boldsymbol{\omega} \in B$ and $\mu_{\widetilde{B}}(\boldsymbol{\omega}) = 0$ for all $\boldsymbol{\omega} \notin B$. As a special case, let $\Omega = \mathbb{R}$ and $B = (a,b)$ be an interval on the real line with $a < b$. Then, B can be written in terms of a fuzzy set as $\widetilde{B} = \{(\omega, 1) : \forall\, \omega \in B\} \cup \{(\omega, 0) : \forall\, \omega \notin B\}$.

It is important to stress that membership and probability density functions are intrinsically different. For example, if $\pi(\boldsymbol{\omega})$ is a density function, *i.e.*, $\pi(\boldsymbol{\omega}) \geq 0$ for all $\boldsymbol{\omega} \in \Omega$ and $\int_\Omega \pi(\boldsymbol{\omega}) d\boldsymbol{\omega} = 1$, then we can obtain a membership function by defining $\mu_{\widetilde{A}}(\boldsymbol{\omega}) = C^{-1}\pi(\boldsymbol{\omega})$, provided that $C = \sup_{\omega \in \Omega} \pi(\boldsymbol{\omega}) < \infty$. However, the converse is not necessarily true, since a membership function does not need to be integrable over Ω.

For probability density functions, it is common to define a support to characterize the set of all points with positive density. For membership functions, we have the same definition to represent the set of all points with positive membership in the fuzzy set. Definition 2 formalizes this concept.

Definition 2. The support of a fuzzy set $\widetilde{A}$ is defined as

$$supp(\widetilde{A}) = \left\{ \boldsymbol{\omega} \in \Omega :\ \mu_{\widetilde{A}}(\boldsymbol{\omega}) > 0 \right\}.$$

Notice that, an element $\boldsymbol{\omega}$ has full membership in its respective fuzzy set when its membership is one. In this context, the element $\boldsymbol{\omega}$ fully contains all features required by the fuzzy set. Definition 3

formalizes the set of all points with full membership, that is, all points where their membership functions are equal to one.

Definition 3. The core of a fuzzy set $\widetilde{A}$ is defined as $\text{core}(\widetilde{A}) = \{\boldsymbol{\omega} \in \Omega : \mu_{\widetilde{A}}(\boldsymbol{\omega}) = 1\}$.

When the core has as least one element, we have a normal fuzzy set (see Definition 4).

Definition 4. A fuzzy set $\widetilde{A}$ is called normal if its core is nonempty. In other words, there is at least one point $\boldsymbol{\omega} \in \mathbb{R}^k$ with $\mu_{\widetilde{A}}(\boldsymbol{\omega}) = 1$.

Let $\widetilde{A}$ be a normal fuzzy set. Then, the closer $\mu_{\widetilde{A}}(\boldsymbol{\omega}_0)$ is to one, the more we believe that $\boldsymbol{\omega}_0$ lies in $\text{core}(\widetilde{A})$ and the closer $\mu_{\widetilde{A}}(\boldsymbol{\omega}_0)$ is to zero, the more we believe that $\boldsymbol{\omega}_0$ is not in $\text{core}(\widetilde{A})$. That is, the degree of membership for an element can also be seen as a measure of uncertainty [29].

Let $\widetilde{A}$ and $\widetilde{B}$ be two fuzzy sets with membership functions $\mu_{\widetilde{A}}(\boldsymbol{\omega})$ and $\mu_{\widetilde{B}}(\boldsymbol{\omega})$, respectively. According to Zadeh [5], (see also [31]), if $\Omega \subseteq \mathbb{R}^k$ the common operations are defined as follows:

1. $\widetilde{A} \subseteq \widetilde{B} \Longleftrightarrow \mu_{\widetilde{A}}(\boldsymbol{\omega}) \leq \mu_{\widetilde{B}}(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega$.
2. $\widetilde{A} \equiv \widetilde{B} \Longleftrightarrow \mu_{\widetilde{A}}(\boldsymbol{\omega}) = \mu_{\widetilde{B}}(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega$.
3. $\widetilde{A}^c$ is the complement of $\widetilde{A} \Longleftrightarrow \mu_{\widetilde{A}^c}(\boldsymbol{\omega}) = 1 - \mu_{\widetilde{A}}(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega$.
4. $\widetilde{C} = \widetilde{A} \cup \widetilde{B} \Longleftrightarrow \mu_{\widetilde{C}}(\boldsymbol{\omega}) = \max\{\mu_{\widetilde{A}}(\boldsymbol{\omega}), \mu_{\widetilde{B}}(\boldsymbol{\omega})\}$ for all $\boldsymbol{\omega} \in \Omega$.
5. $\widetilde{D} = \widetilde{A} \cap \widetilde{B} \Longleftrightarrow \mu_{\widetilde{D}}(\boldsymbol{\omega}) = \min\{\mu_{\widetilde{A}}(\boldsymbol{\omega}), \mu_{\widetilde{B}}(\boldsymbol{\omega})\}$ for all $\boldsymbol{\omega} \in \Omega$.

From the above definitions, if we consider $\widetilde{\Omega} = \{(\boldsymbol{\omega}, 1) ; \boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^k\}$ as the universal fuzzy set, then for any fuzzy set $\widetilde{A}$ we have $\widetilde{A} \subseteq \widetilde{\Omega}$, provided that the membership function of $\widetilde{A}$ has domain Ω. In addition, if there exists $\boldsymbol{\omega}_0 \in \Omega$ such that $\max\{\mu_{\widetilde{A}}(\boldsymbol{\omega}_0), \mu_{\widetilde{A}^c}(\boldsymbol{\omega}_0)\} < 1$ we have that $\widetilde{A} \cup \widetilde{A}^c \neq \widetilde{\Omega}$. Also, if there exists $\boldsymbol{\omega}_0 \in \Omega$ such that $\min\{\mu_{\widetilde{A}}(\boldsymbol{\omega}_0), \mu_{\widetilde{A}^c}(\boldsymbol{\omega})\} > 0$, we have that $\widetilde{A} \cap \widetilde{A}^c \neq \widetilde{\varnothing}$. As the reader can see, these properties block the excluded middle and contradiction laws of classic set theory (for further details see [31–33]).

The concept of a fuzzy set is very broad and difficult to handle without some additional specifications. In this context, the next definition allows us to specify fuzzy numbers, which are natural extensions of traditional numbers. However, this latter definition depends on fuzzy convexity, which is defined next (see [5] for more details).

Definition 5. A set $\widetilde{A}$ is convex if and only if $\mu_{\widetilde{A}}(\lambda\boldsymbol{\omega}_1 + (1-\lambda)\boldsymbol{\omega}_2) \geq \min\{\mu_{\widetilde{A}}(\boldsymbol{\omega}_1), \mu_{\widetilde{A}}(\boldsymbol{\omega}_2)\}$ for all $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \in \Omega$ and $\lambda \in [0, 1]$.

Note that the concept of convexity under the fuzzy approach differs from the classic definition of convexity under functional analysis. More discussion about this concept will be presented in Section 3. A fuzzy interval $\widetilde{A}$ is a fuzzy set that satisfies the condition of convexity and normality, so the core of a fuzzy interval is constituted by all elements with membership one. A fuzzy interval $\widetilde{A}$ is a fuzzy number when the cardinality of $core(\widetilde{A})$ equals 1 [28]. Fuzzy numbers and fuzzy intervals are useful to represent imprecision for point and interval measures, respectively. As mentioned earlier, these concepts have multiple applications, e.g., in artificial intelligence, image processing, speech recognition, biological, and medical science, operations research, decision analysis, information processing, economics, geography, psychology, linguistics, *etc*. More applications can be found in [34,35]. Figure 1a–c illustrate a general fuzzy set, a fuzzy interval and a fuzzy number, respectively.

Dubois *et al*. [27] defined the class of LR (left and right) membership functions defined over $\Omega = \mathbb{R}$, *i.e.*, the class of membership functions that can be entirely characterized by three parameters, namely, (m, α, β), and two functions L and R. The next definition is related to the concept of LR-type fuzzy numbers.

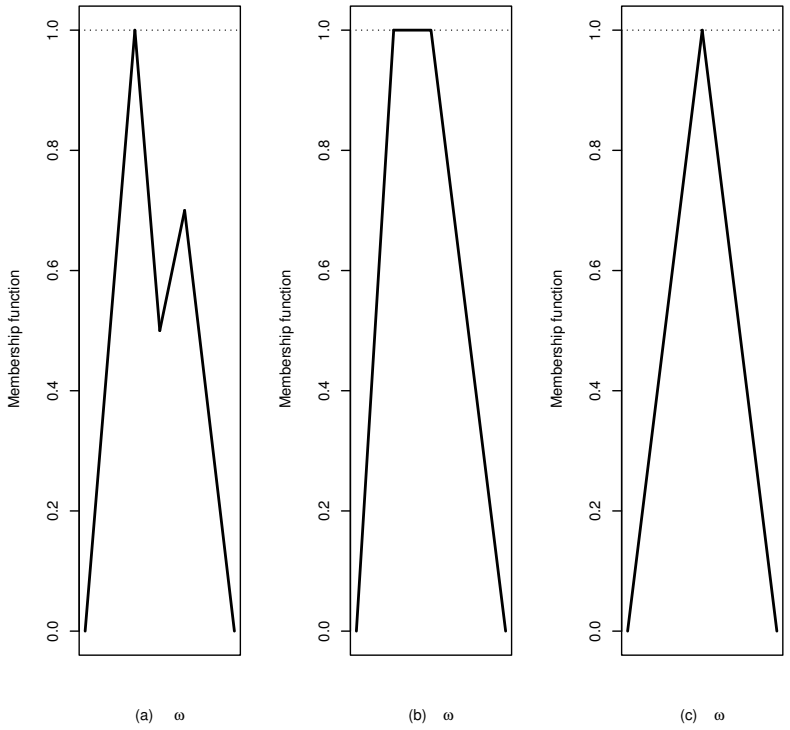


Figure 1. Representations of (**a**) a general fuzzy set, (**b**) a fuzzy interval and (**c**) a fuzzy number.

Definition 6. The fuzzy number $\widetilde{A}$ is said to be of LR-type if two decreasing functions exist $L, R : [0, +\infty) \to [0, 1]$ with $L(0) = R(0) = 1$, $\lim_{\omega \to +\infty} L(\omega) = \lim_{\omega \to +\infty} R(\omega) = 0$ and positive real numbers $m \geq 0$, $\alpha > 0$, $\beta > 0$ such that

$$\mu_{\widetilde{A}}(\omega) = \begin{cases} L\left(\dfrac{m-\omega}{\alpha}\right), & \text{for } \omega \leq m, \\ R\left(\dfrac{\omega-m}{\beta}\right), & \text{for } \omega \geq m, \end{cases}$$

where m is called the center of $\widetilde{A}$ and α and β are called the left and right propagations, respectively.

If $\alpha = \beta$, $\widetilde{A}$ is called a symmetric fuzzy number. For a symmetric membership function, the equality $L\left(\frac{m-\omega}{\alpha}\right) = R\left(\frac{\omega-m}{\beta}\right)$ holds for all $\omega \in \mathbb{R}$.

In this paper, we use all definitions presented in this section to connect the classic statistical quantities with fuzzy theory.

3. Confidence Sets and Membership Functions

As mentioned in the Introduction, the main goal of this paper is to infer about some parameter $\boldsymbol{\theta}$ using the confidence sets under the fuzzy set theory. For that reason, we start this section by presenting the definition of a general confidence set.

Under a parametric statistical model $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ with $\mathcal{P} = \{P^{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$ and $k < \infty$, a $(1-\alpha)$ confidence set is a function $C_\alpha : \mathcal{X} \to 2^\Theta$, where 2^Θ is the family of all subsets of Θ (the power set) satisfying

$$P^{\boldsymbol{\theta}}(C_\alpha(\mathbf{X}) = \varnothing) = 0, \quad P^{\boldsymbol{\theta}}(C_\alpha(\mathbf{X}) \ni \boldsymbol{\theta}) \geq 1-\alpha \quad \text{and} \quad \inf_{\boldsymbol{\theta} \in \Theta} P^{\boldsymbol{\theta}}(C_\alpha(\mathbf{X}) \ni \boldsymbol{\theta}) = 1-\alpha \tag{1}$$

for every $\boldsymbol{\theta} \in \Theta$, where $\mathbf{X} \in \mathcal{X}$ is a random vector defined in the statistical model (see [36], p. 315). When $P^{\boldsymbol{\theta}}(C_\alpha(\mathbf{X}) \ni \boldsymbol{\theta}) = 1-\alpha$ for all $\boldsymbol{\theta} \in \Theta$, then the confidence set C_α is exact. Procedures to build confidence sets can be found in [37–41] among others. These procedures are in general based on pivotal quantities and likelihood-ratio statistics. Here, $1-\alpha$ is called the confidence level and α is called the significance level [42]. Intuitively, the interval width depends on the confidence level, for instance, the greater the confidence level the greater the interval width built under normal distributions (see [43]).

After observing the sample, the confidence set $C_\alpha(\mathbf{x})$ is a fixed set and $P^{\boldsymbol{\theta}}(C_\alpha(\mathbf{x}) \ni \boldsymbol{\theta})$ is zero or one, where $\mathbf{x}$ is the observed sample, so the probability statements in Equation (1) are used just to construct a proper confidence set. Once the sample is observed, this confidence set is fixed. Therefore, the observed confidence sets cannot be interpreted in terms of probabilities (see [44] Section 3.1.2, p. 41, for further details). In this section we show that, although it is not possible to make probabilistic statements about observed confidence sets, we can interpret the observed confidence sets in terms of fuzzy sets. We present a general membership function that provides all information contained in an observed confidence set $C_\alpha(\mathbf{x})$, for all levels $\alpha \in [0,1]$ (alpha-cuts).

Lemma 1. *Let C_α be a confidence set of $\alpha \in [0,1]$ level for some subset of Θ, where $\mathbf{x} \in \mathcal{X}$. Let $\mathcal{C} = \{C_\alpha(\mathbf{x})\}_{\alpha \in I}$ where $I \subseteq [0,1]$ is non-empty. For each $\boldsymbol{\theta} \in \Theta$, we define*

$$\mu_{\widetilde{\Theta},\mathcal{C}}(\boldsymbol{\theta}) = \max\{0, \sup\{\alpha \in I:\ C_\alpha(\mathbf{x}) \cap \{\boldsymbol{\theta}\} \neq \varnothing\}\}, \tag{2}$$

where $\sup\{\varnothing\} = -\infty$. We use the short notation $\mu_{\widetilde{\Theta}}$ when the family $\mathcal{C}$ is not the focus. Then, $\mu_{\widetilde{\Theta}}$ is a membership function, that is, $\mu_{\widetilde{\Theta}}: \Theta \to [0,1]$.

Proof. As $\mu_{\widetilde{\Theta},\mathcal{C}}$ is a function and $I \subseteq [0,1]$ is non-empty. The proof is straightforward. □

For each proposed confidence region, we can represent the parameter space by the fuzzy set

$$\widetilde{\Theta} = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}));\ \boldsymbol{\theta} \in \Theta\},$$

where $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta})$ is given in Equation (2). Note that for different confidence sets $C_{1,\alpha}$ and $C_{2,\alpha}$ we have different memberships, namely $\mu_{\widetilde{\Theta},\mathcal{C}_1} \equiv \mu_{\widetilde{\Theta}_1}$ and $\mu_{\widetilde{\Theta},\mathcal{C}_2} \equiv \mu_{\widetilde{\Theta}_2}$, respectively, where $\mathcal{C}_1 = \{C_{1,\alpha}(\mathbf{x})\}_{\alpha \in I_1}$ and $\mathcal{C}_2 = \{C_{2,\alpha}(\mathbf{x})\}_{\alpha \in I_2}$ with $I_1 \cup I_2 \subset [0,1]$. As a consequence, the resulting fuzzy sets will have different representations, namely

$$\widetilde{\Theta}_1 = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}_1}(\boldsymbol{\theta}));\ \boldsymbol{\theta} \in \Theta\} \quad \text{and} \quad \widetilde{\Theta}_2 = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}_2}(\boldsymbol{\theta}));\ \boldsymbol{\theta} \in \Theta\},$$

respectively. Additional information with respect to Equation (2) can be found in Mauris *et al.* [45] and more recently in [11]. Patriota [11] studied some relationships with p-values when the confidence region C_α is built under the likelihood-ratio statistic.

The next result characterizes the core of $\widetilde{\Theta}$.

Theorem 1. *Let $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta})$ and $I = [0,1]$ be the quantities defined in Equation (2), then core$(\widetilde{\Theta}) = \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} \in C_\alpha(\mathbf{x})\ \forall \alpha \in I\}$.*

Proof. From Definition 3, we have that $\text{core}(\widetilde{\Theta}) = \{\boldsymbol{\theta} \in \Theta : \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1\}$. Now, we have that $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1$ if, and only if, $\sup\{\alpha \in I : C_\alpha(\mathbf{x}) \cap \{\boldsymbol{\theta}\} \neq \varnothing\} = 1$. This last fact occurs when $\boldsymbol{\theta} \in C_\alpha(\mathbf{x})$ for all $\alpha \in I$ concluding the proof. □

Theorem 1 characterizes the functional form of the core for the membership function defined in Equation (2), that is, those values in Θ that produce full membership.

Remark 1. If the confidence set $C_\alpha(\mathbf{x})$ is centered in the maximum likelihood (ML) estimate $\widehat{\boldsymbol{\theta}}$, then we have that $\mu_{\widetilde{\Theta}}(\widehat{\boldsymbol{\theta}}) = 1$. This means that the ML estimate is part of the core of the fuzzy set $\widetilde{\Theta}$. We can interpret core$(\widetilde{\Theta})$ as the set of all parameter values for which the related probability distribution explains the observed data according to $\mathcal{C}$.

Next we define non-increasing confidence sets in terms of the significance level α.

Definition 7. Let $\mathcal{C} = \{C_\alpha(\mathbf{x})\}_{\alpha \in I}$ be a family of confidence sets. We say that $\mathcal{C}$ is a non-increasing family of confidence sets if $C_{\alpha_1}(\mathbf{x}) \subseteq C_{\alpha_2}(\mathbf{x})$ for all $0 \leq \alpha_2 < \alpha_1 \leq 1$.

The next result relates the monotonicity property of confidence sets with the membership functions.

Theorem 2. *Let $\mathcal{C} = \{C_\alpha(\mathbf{x})\}_{\alpha \in I}$ be a family of non-increasing confidence sets. If core$(\widetilde{\Theta}) = \{\boldsymbol{\theta}^*\}$, then $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) \geq \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}^+)$, for all $\boldsymbol{\theta} \in \theta_\epsilon$ and $\boldsymbol{\theta}^+ \in (\theta_\gamma \cap \theta_\epsilon^c)$, where $\theta_\nu = \{\boldsymbol{\theta} \in \Theta; \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \nu\}$ and with $0 < \epsilon \leq \gamma$.*

Proof. As $\theta_\epsilon \subseteq \theta_\gamma$ and $C_\alpha(\mathbf{x})$ is non-increasing, we have for each $\boldsymbol{\theta} \in \theta_\epsilon$ and $\boldsymbol{\theta}^+ \in (\theta_\gamma \cap \theta_\epsilon^c)$ that

$$\{\alpha \in I : C_\alpha(\mathbf{x}) \cap \{\boldsymbol{\theta}^+\} \neq \varnothing\} \subseteq \{\alpha \in I : C_\alpha(\mathbf{x}) \cap \{\boldsymbol{\theta}\} \neq \varnothing\}.$$

Applying the supremum, we obtain $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) \geq \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}^+)$, concluding the proof. □

Another important result for non-increasing confidence sets is the relationship between the convexity concept in Definition 5, defined by Zadeh [5], and the monotonicity property. This relationship is based on Theorem 2, concluding that $\mu_{\widetilde{\Theta}}(\lambda\boldsymbol{\theta}^+ + (1-\lambda)\boldsymbol{\theta}) \geq \min\{\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}^+), \mu_{\widetilde{\Theta}}(\boldsymbol{\theta})\}$, for all $\boldsymbol{\theta}^+, \boldsymbol{\theta} \in \Theta$ and $\lambda \in [0,1]$ and therefore $\widetilde{\Theta}$ is convex in the fuzzy context. Theorem 2 can be used to compare the membership functions. Next definition introduces the concept of supremacy for comparing confident sets.

Definition 8. Let $\mu_{\widetilde{\Theta}_1}, \mu_{\widetilde{\Theta}_2} : \Theta \to [0,1]$ be two membership functions with the same core. We say that $\mu_{\widetilde{\Theta}_1}$ has total supremacy over $\mu_{\widetilde{\Theta}_2}$ if $\mu_{\widetilde{\Theta}_1}(\boldsymbol{\theta}) \leq \mu_{\widetilde{\Theta}_2}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$. They are totally equivalent if $\mu_{\widetilde{\Theta}_1}(\boldsymbol{\theta}) = \mu_{\widetilde{\Theta}_2}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$. We denote total supremacy by $\mu_{\widetilde{\Theta}_1} \ll_T \mu_{\widetilde{\Theta}_2}$ and total equivalence by $\mu_{\widetilde{\Theta}_1} \sim_T \mu_{\widetilde{\Theta}_2}$.

Definition 8 allows us to compare two different confidence sets, e.g., we can determine if a confidence set is more conservative than another for all confidence levels. Definition 8 is similar to the definition of *superiority* given by Xie and Singh [46]. Note also that if $\mathcal{U}$ is the family of all membership functions then $\ll_T$ establishes an order relation in $\mathcal{U}$ and this relation is the order relation of the inclusion for fuzzy sets, (see for instance [47]). Notice that Definition 8 is strong and is not applicable in many situations, notably if two membership functions have different core sets. In order to make the supremacy concept less restrictive, allowing us to include more situations, we define the following operators.

Definition 9. Let $\mu_{\widetilde{\Theta}_1}$ and $\mu_{\widetilde{\Theta}_2}$ two continuous membership functions. We say that

1. $\mu_{\widetilde{\Theta}_1}$ has r-up-supremacy over $\mu_{\widetilde{\Theta}_2}$ if $E^r(\mu_{\widetilde{\Theta}_1}) \leq E^r(\mu_{\widetilde{\Theta}_2})$, and is denoted by $\mu_{\widetilde{\Theta}_1} \ll_{E^r} \mu_{\widetilde{\Theta}_2}$. Moreover, they are r-up-equivalent when $E^r(\mu_{\widetilde{\Theta}_1}) = E^r(\mu_{\widetilde{\Theta}_2})$, and are denoted by $\mu_{\widetilde{\Theta}_1} \sim_{E^r} \mu_{\widetilde{\Theta}_2}$, where
$$E^r(\mu_{\widetilde{\Theta}}) = \int_{\Theta_r} \mu_{\widetilde{\Theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$
with
$$\Theta_r = \{\boldsymbol{\theta} \in \Theta : \ \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) \geq r\}$$

and $r \in [0,1]$ is such that the resulting integral is well defined.

2. If $\mu_{\widetilde{\Theta}}$ is integrable with respect to the Lebesgue measure in dimension $k = \dim(\Theta)$, $\mu_{\widetilde{\Theta}_1}$ has r-down-supremacy over $\mu_{\widetilde{\Theta}_2}$ if $E_r(\mu_{\widetilde{\Theta}_1}) \leq E_r(\mu_{\widetilde{\Theta}_2})$, denoted by $\mu_{\widetilde{\Theta}_1} \ll_{E_r} \mu_{\widetilde{\Theta}_2}$. Moreover, they are r-down-equivalent when $E_r(\mu_{\widetilde{\Theta}_1}) = E_r(\mu_{\widetilde{\Theta}_2})$, and it is denoted by $\mu_{\widetilde{\Theta}_1} \sim_{E_r} \mu_{\widetilde{\Theta}_2}$, where

$$E_r(\mu_{\widetilde{\Theta}}) = \int_{\Theta_r^c} \mu_{\widetilde{\Theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

with

$$\Theta_r^c = \{\boldsymbol{\theta} \in \Theta : \ \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) < r\}.$$

We call E^r the r-up integral operator and E_r the r-down integral operator.

Notice that, if $\mu_{\widetilde{\Theta}}$ is integrable, then by Definition 9, it is straightforward that

$$E^0(\mu_{\widetilde{\Theta}}) = E_r(\mu_{\widetilde{\Theta}}) + E^r(\mu_{\widetilde{\Theta}}) = E_1(\mu_{\widetilde{\Theta}}) \tag{3}$$

for all $r \in [0,1]$. In addition, if $\dim(\{\boldsymbol{\theta} \in \Theta : \ \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1\}) < k$, we also have that $E^1(\mu_{\widetilde{\Theta}}) = E_0(\mu_{\widetilde{\Theta}}) = 0$. Example 1 describes how to analytically compute the quantities E^r and E_r. However, for more complex models, analytical solutions are virtually impossible, so these integrals have to be computed numerically by using any software (for instance, `MAPLE`, `MATLAB`, `Ox`, `R`, `SAS`).

Example 1. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$ be a random sample (the random variables are independent and identically distributed) from a normal population with mean θ and known variance σ^2 and let $\mathbf{x}$ be the observed sample. Here, $\Theta = \mathbb{R}$ and a $(1-\alpha)$ confidence set for θ, using the pivotal quantity method, is given by

$$C_\alpha(\mathbf{x}) = \left[\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\right],$$

for $\alpha \in I = [0,1]$, where $\bar{x}$ is the sample mean and z_q is the qth-quantile of a standard normal distribution. The membership function is given by

$$\mu_{\widetilde{\Theta}}(\theta) = 2\left(1 - \Phi\left(\frac{\sqrt{n}|\bar{x} - \theta|}{\sigma}\right)\right).$$

Then, solving the equations $\mu_{\widetilde{\Theta}}(c_{1r}) = \mu_{\widetilde{\Theta}}(c_{2r}) = r$ with respect to c_{1r} and c_{2r}, where $c_{1r} \leq \bar{x} \leq c_{2r}$, we obtain

$$c_{1r} = \bar{x} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{r}{2}\right), \quad c_{2r} = \bar{x} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{r}{2}\right)$$

and

$$E_r(\mu_{\widetilde{\Theta}}) = \int_{\Theta_r^c} \mu_{\widetilde{\Theta}}(\theta)\, d\theta = 4\int_{-\infty}^{c_{1r}} \Phi\left(\frac{\sqrt{n}(\theta - \bar{x})}{\sigma}\right) d\theta,$$

where $\Theta_r = \mu_{\widetilde{\Theta}}^{-1}([r,1]) = [c_{1r}, c_{2r}]$. Letting $u = \sqrt{n}(\theta - \bar{x})/\sigma$, we have that

$$\int_{-\infty}^{c_{1r}} \Phi\left(\frac{\sqrt{n}(\theta - \bar{x})}{\sigma}\right) d\theta = \frac{\sigma}{\sqrt{n}}\int_{-\infty}^{\sqrt{n}(c_{1r} - \bar{x})/\sigma} \Phi(u)\, du.$$

Now, by using the identity [48] $\int_{-\infty}^{a} \Phi(x)\, dx = a\Phi(a) + \phi(a)$, where $\phi(\cdot)$ stands for the standard normal density function, we have that

$$E_r(\mu_{\widetilde{\Theta}}) = \frac{4\sigma}{\sqrt{n}}\left\{\frac{\sqrt{n}(c_{1r} - \bar{x})}{\sigma}\Phi\left(\frac{\sqrt{n}(c_{1r} - \bar{x})}{\sigma}\right) + \phi\left(\frac{\sqrt{n}(c_{1r} - \bar{x})}{\sigma}\right)\right\}.$$

Similarly for the r-up integral operator,

$$E^r(\mu_{\widetilde{\Theta}}) = 4\int_{-\infty}^{\bar{x}} \Phi\left(\frac{\sqrt{n}(\theta-\bar{x})}{\sigma}\right) d\theta - 4\int_{-\infty}^{c_{1r}} \Phi\left(\frac{\sqrt{n}(\theta-\bar{x})}{\sigma}\right) d\theta,$$

then

$$E^r(\mu_{\widetilde{\Theta}}) = \frac{4\sigma}{\sqrt{n}}\left\{\phi(0) - \frac{\sqrt{n}(c_{1r}-\bar{x})}{\sigma}\Phi\left(\frac{\sqrt{n}(c_{1r}-\bar{x})}{\sigma}\right) - \phi\left(\frac{\sqrt{n}(c_{1r}-\bar{x})}{\sigma}\right)\right\}.$$

Definition 9 will be used to identify more conservative confidence sets. It is possible to show that $\ll_{E_r}$, $\ll_{E^r}$ and $\ll_T$ satisfy the requirements to be order relations (reflexivity, antisymmetry and transitivity) and $\sim_{E_r}$, $\sim_{E^r}$ and $\sim_T$ satisfy the requirements to be equivalence relations (reflexivity, symmetry and transitivity).

Theorem 3 establishes some relations among the three types of supremacies.

Theorem 3. *Let $\mu_{\widetilde{\Theta}_1}, \mu_{\widetilde{\Theta}_2} : \Theta \to [0,1]$ be two continuous and integrable membership functions. Then,*

$$[\forall r \in [0,1],\ \ \mu_{\widetilde{\Theta}_1} \ll_{E_r} \mu_{\widetilde{\Theta}_2}] \text{ if and only if } [\forall r \in [0,1],\ \ \mu_{\widetilde{\Theta}_1} \ll_{E^r} \mu_{\widetilde{\Theta}_2}]. \tag{4}$$

Moreover, if $core(\widetilde{\Theta}_1) = core(\widetilde{\Theta}_2)$,

$$[\forall r \in [0,1],\ \ \mu_{\widetilde{\Theta}_1} \ll_{E^r} \mu_{\widetilde{\Theta}_2}] \text{ if and only if } [\mu_{\widetilde{\Theta}_1} \ll_T \mu_{\widetilde{\Theta}_2}], \tag{5}$$

where $\widetilde{\Theta}_i = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}_i}) :\ \boldsymbol{\theta} \in \Theta\}$.

Proof. Let $\boldsymbol{\theta}^* \in \Theta$ and $k \in \mathbb{R}$ and define $\mu_{\widetilde{\Theta}_{1,k,*}}(\boldsymbol{\theta}) = \mu_{\widetilde{\Theta}_1}(\boldsymbol{\theta} + k\boldsymbol{\theta}^*)$. Notice that $E_r(\mu_{\widetilde{\Theta}_1}) \leq E_r(\mu_{\widetilde{\Theta}_2})$, for all $r \in [0,1]$, if and only if there exist $\boldsymbol{\theta}^* \in \Theta$ and $k \in \mathbb{R}$ such that $\widetilde{\Theta}_{1,k,*} \subseteq \widetilde{\Theta}_2$, where the latter is the inclusion of fuzzy sets and $\widetilde{\Theta}_{1,k,*}$ depends on the membership function $\mu_{\widetilde{\Theta}_{1,k,*}}$. This implies that $E^r(\mu_{\widetilde{\Theta}_1}) \leq E^r(\mu_{\widetilde{\Theta}_2})$ for all $r \in [0,1]$. The proof of the converse is similar. If $k = 0$, by the equality of the cores, we have $\widetilde{\Theta}_1 \subseteq \widetilde{\Theta}_2$ if and only if $\mu_{\widetilde{\Theta}_1} \ll_T \mu_{\widetilde{\Theta}_2}$. □

A confidence interval is said to be more conservative than another if the former interval's amplitude is greater than the latter's for a specific significance level. A procedure to generate a confidence interval is considered more conservative than another if the interval's amplitude is greater than the latter's for all significance levels. Below, we define the conservative concept for general confidence sets.

Definition 10. Let $\mathcal{C}_1 = \{C_{1,\alpha}(\mathbf{x})\}_{\alpha \in I_1}$ and $\mathcal{C}_2 = \{C_{2,\alpha}(\mathbf{x})\}_{\alpha \in I_2}$ be two families of confidence sets. We say that $\mathcal{C}_2$ is more conservative than $\mathcal{C}_1$ if their respective memberships $\mu_{\widetilde{\Theta},\mathcal{C}_2} \equiv \mu_{\widetilde{\Theta}_2}$ and $\mu_{\widetilde{\Theta},\mathcal{C}_1} \equiv \mu_{\widetilde{\Theta}_1}$ satisfy $\mu_{\widetilde{\Theta}_1} \ll_{E_r} \mu_{\widetilde{\Theta}_2}$ for all $r \in [0,1]$. We say that the region $C_{2,\alpha}(\mathbf{x})$ is up-more conservative than $C_{1,\alpha}(\mathbf{x})$ if $\mu_{\widetilde{\Theta}_1} \ll_{E^\alpha} \mu_{\widetilde{\Theta}_2}$ and down-more conservative if $\mu_{\widetilde{\Theta}_1} \ll_{E_\alpha} \mu_{\widetilde{\Theta}_2}$.

The next proposition establishes some properties of $\widetilde{\Theta}$. The proof is straightforward from Lemma 1, Definition 3 and Definition 8 and therefore is omitted.

Proposition 1. *Let $\widetilde{\Theta} = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}));\ \boldsymbol{\theta} \in \Theta\}$ with $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta})$ as in Equation (2). Then, the following statements are valid.*

1. *$core(\widetilde{\Theta}) = \mu^{-1}_{\widetilde{\Theta}}(\{1\})$, where $\mu^{-1}_{\widetilde{\Theta}}(\{1\}) = \{\boldsymbol{\theta} \in \Theta :\ \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1\}$. Also, if $core(\widetilde{\Theta})$ is a singleton then $\widetilde{\Theta}$ is a fuzzy number.*
2. *If the cardinality of $\mu^{-1}_{\widetilde{\Theta}}(\{1\})$ is greater than 1, then $\widetilde{\Theta}$ is a normal fuzzy set.*
3. *If $\mu^{-1}_{\widetilde{\Theta}}(\{1\}) = \Theta$, then $\widetilde{\Theta} = \{(\boldsymbol{\theta}, 1);\ \boldsymbol{\theta} \in \Theta\}$.*

4. *Let $\widetilde{\Theta}_1$ and $\widetilde{\Theta}_2$ be two fuzzy set representations over the parametric space Θ associated with $C_{1,\alpha}(\mathbf{x})$ and $C_{2,\alpha}(\mathbf{x})$ respectively. If $C_{1,\alpha}(\mathbf{x}) \subseteq C_{2,\alpha}(\mathbf{x})$ for all $\alpha \in [0,1]$, then $\mu_{\widetilde{\Theta}_2} \ll \mu_{\widetilde{\Theta}_1}$ and $C_{2,\alpha}(\mathbf{x})$ is more conservative than $C_{1,\alpha}(\mathbf{x})$ for all $\alpha \in [0,1]$.*

Definitions 8–10 are tools for comparing confidence sets through their respective membership functions. The membership functions used are defined in the same parameter space. However, there are situations in which we are interested in comparing confidence sets from a partial vector parameter with confidence sets from a full parameter vector. Therefore, a membership function for partial parameter vectors is defined next.

Let $\boldsymbol{\theta} = (\boldsymbol{\lambda}^\top, \boldsymbol{\psi}^\top)^\top$, where $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$ are vectors with dimensions k_1 and k_2 with $k = k_1 + k_2$. Without loss of generality, let $C^*_\alpha(\mathbf{x})$ be a confidence set for $\boldsymbol{\lambda}$ and let Λ be the set in which $\boldsymbol{\lambda}$ varies. Then, a membership function for $\boldsymbol{\lambda}$ can be defined simply by $\mu_{\widetilde{\Lambda}} : \Lambda \to [0,1]$ such that

$$\mu_{\widetilde{\Lambda}}(\boldsymbol{\lambda}) = \max\{0, \sup\{\alpha \in [0,1] : \ C^*_\alpha(\mathbf{x}) \cap \{\boldsymbol{\lambda}\} \neq \varnothing\}\}. \tag{6}$$

The same properties of the membership (2) and the above definitions are valid for this partial membership function.

4. Examples

In this section we present some examples of confidence sets in order to illustrate the relation between the confidence set and associated membership function. Moreover, we compute the *r*-up and down integral operators.

4.1. Exponential Distribution

Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be a random sample from an exponential distribution with rate θ, let $\mathbf{x}$ be the observed sample and $\bar{x}$ the sample mean. Then, a $1-\alpha$ level confidence interval for θ

(a_1) by using the pivotal quantity (see [49], p. 267) is

$$C_{1,\alpha}(\mathbf{x}) = \left[\frac{\chi^2_{2n;\alpha/2}}{2n\bar{x}}; \frac{\chi^2_{2n;1-\alpha/2}}{2n\bar{x}}\right]$$

(a_2) by using the asymptotic approximation [50] is

$$C_{2,\alpha}(\mathbf{x}) = \left[\frac{1}{\bar{x}} - \frac{z_{1-\alpha/2}}{\bar{x}\sqrt{n}}; \frac{1}{\bar{x}} + \frac{z_{1-\alpha/2}}{\bar{x}\sqrt{n}}\right],$$

where $\chi^2_{\nu;q}$ is the q-th quantile of a chi square distribution with ν degrees of freedom.

Thus, from Lemma 1 we have the following:

(a_1) The membership function for $\mathcal{C}_1 = \{C_{1,\alpha}\}_{0\leq\alpha\leq1}$ is $\mu_{\widetilde{\Theta}_1} : \mathbb{R} \to [0,1]$ such that

$$\mu_{\widetilde{\Theta}_1}(\theta) = \begin{cases} 2\,\chi^2\big(2n\theta\bar{x}; 2n\big), & \text{for } \theta < 1/\bar{x}, \\ 2\left(1 - \chi^2\big(2n\theta\bar{x}; 2n\big)\right), & \text{for } \theta \geq 1/\bar{x} \end{cases}$$

(a_2) The membership function for $\mathcal{C}_2 = \{C_{2,\alpha}\}_{0\leq\alpha\leq1}$ is $\mu_{\widetilde{\Theta}_2} : \mathbb{R} \to [0,1]$ such that

$$\mu_{\widetilde{\Theta}_2}(\theta) = 2\,(1 - \Phi(\sqrt{n}|1 - \bar{x}\theta|)),$$

where $\chi^2(\cdot;\nu)$ is the cumulative distribution function of a chi square distribution with ν degrees of freedom.

Figure 2 presents the membership functions (panel a) and functions (r, E_r) for $r \in [0,1]$ and (r, E^r) for $r \in [0,1]$ for both confidence intervals (panels b and c), considering $n = 5$ and $\bar{x} = 1$. In all plots, the solid and dotted lines represent cases (a_1) and (a_2), respectively.

In particular, Figure 2a shows that for some $\theta \in \Theta^* \subset \Theta$ we have $\mu_{\widetilde{\Theta}_1}(\theta) \leq \mu_{\widetilde{\Theta}_2}(\theta)$ and for some $\theta \in \Theta' \subset \Theta$ we have $\mu_{\widetilde{\Theta}_2}(\theta) \leq \mu_{\widetilde{\Theta}_1}(\theta)$. That is, the order relation is not the same for all $\theta \in \Theta$. However, Definition 8 states that the order relation must be fulfilled for all $\theta \in \Theta$. In Figure 2b, it can be observed that the inequality $E_r(\mu_{\widetilde{\Theta}_1}) \leq E_r(\mu_{\widetilde{\Theta}_2})$ holds for $r \in [0.81; 1.00]$. Consequently, $\mu_{\widetilde{\Theta}_1}$ has r-down-supremacy over $\mu_{\widetilde{\Theta}_2}$ for all $r \in [0.81; 1]$ (see Definition 9). For $r = 0.81$, we have $E_r(\mu_{\widetilde{\Theta}_1}) = E_r(\mu_{\widetilde{\Theta}_2}) = 0.55$, while for values of $r \leq 0.81$ we observe that $E_r(\mu_{\widetilde{\Theta}_1})$ and $E_r(\mu_{\widetilde{\Theta}_2})$ have interlacings. Finally, Figure 2c shows that the relationship $E^r(\mu_{\widetilde{\Theta}_2}) \leq E^r(\mu_{\widetilde{\Theta}_1})$ is true for $r \in [0.40; 0.84]$, then we observe that $\mu_{\widetilde{\Theta}_2}$ has r-up-supremacy over $\mu_{\widetilde{\Theta}_1}$ for all $r \in [0.4; 0.84]$. For $r = 0.4$ and $r = 0.84$, we have $E^r(\mu_{\widetilde{\Theta}_1}) = E^r(\mu_{\widetilde{\Theta}_2}) = 0.41$ and $E^r(\mu_{\widetilde{\Theta}_1}) = E^r(\mu_{\widetilde{\Theta}_2}) = 0.14$, respectively. For $r \geq 0.84$ we observe that $\mu_{\widetilde{\Theta}_1}$ has r-up-supremacy over $\mu_{\widetilde{\Theta}_2}$.

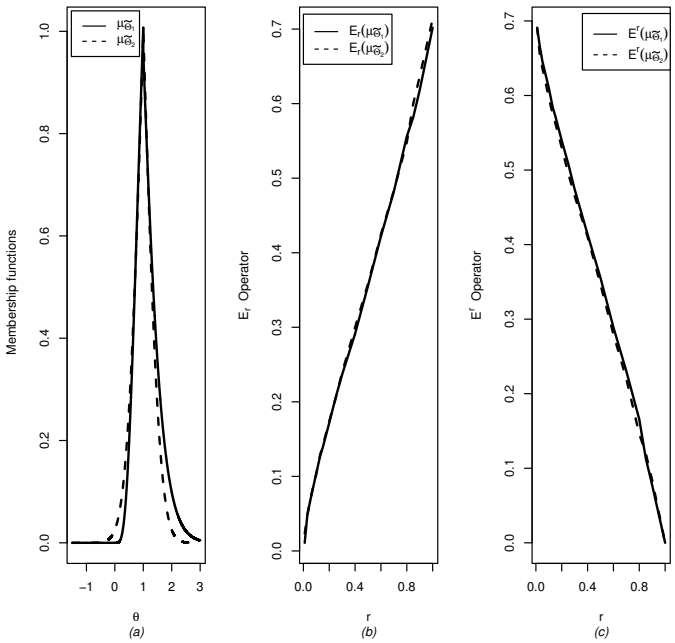


Figure 2. (**a**) membership functions for the exponential rate based on the pivotal quantity (solid line) and based on the normal approximation (dotted line); (**b**) the r-down integral operator for the respective memberships; (**c**) the r-up integral operator for the respective memberships. $n = 5$ and $\bar{x} = 1$ are considered.

4.2. Poisson Distribution

Now, let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ a random sample from a Poisson distribution with rate θ, and let $\mathbf{x}$ be the observed sample and $\bar{x}$ the sample mean. A $1 - \alpha$ level confidence interval for θ

(a_1) by using the pivotal quantity [51] is

$$C_{1,\alpha}(\mathbf{x}) = \left[\frac{\chi^2_{2n\bar{x};\alpha/2}}{2n}; \frac{\chi^2_{2n\bar{x}+2;1-\alpha/2}}{2n} \right]$$

(a_2) by using the asymptotic approximation [52] is

$$C_{2,\alpha}(\mathbf{x}) = \left[\bar{x} - \frac{z_{1-\alpha/2}\sqrt{\bar{x}}}{\sqrt{n}}; \bar{x} + \frac{z_{1-\alpha/2}\sqrt{\bar{x}}}{\sqrt{n}} \right].$$

Thus, from Lemma 1 we have the following.

(a_1) The membership function for $\mathcal{C}_1 = \{C_{1,\alpha}\}_{0\leq\alpha\leq 1}$ is $\mu_{\widetilde{\Theta}_1} : \mathbb{R} \to [0,1]$ such that

$$\mu_{\widetilde{\Theta}_1}(\theta) = \begin{cases} \min\left\{1, 2\left(1 - \chi^2(2n\theta; 2n\bar{x})\right)\right\}, & \text{for } \theta \leq \bar{x}, \\ \min\left\{1, 2\,\chi^2(2n\theta; 2n\bar{x} + 2)\right\}, & \text{for } \theta \geq \bar{x} \end{cases}$$

(a_2) The membership function for $\mathcal{C}_2 = \{C_{2,\alpha}\}_{0\leq\alpha\leq 1}$ is $\mu_{\widetilde{\Theta}_2} : \mathbb{R} \to [0,1]$ such that

$$\mu_{\widetilde{\Theta}_2}(\theta) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\,|\bar{x} - \theta|}{\sqrt{\bar{x}}}\right)\right)$$

Figure 3a presents the membership functions. Note that in this case, the total supremacy does not occur (see Definition 8). Moreover, Figure 3b,c present the functions (r, E_r) and (r, E^r) for $r \in [0,1]$ for both confidence intervals respectively. From these figures, it can be concluded that $\mu_{\widetilde{\Theta}_1}$ has r-up-supremacy over $\mu_{\widetilde{\Theta}_2}$ for all $r \in [0,1]$. We considered $n = 5$ and $\bar{x} = 4$, and as in the previous example, the solid and dotted lines represent cases (a_1) and (a_2), respectively.

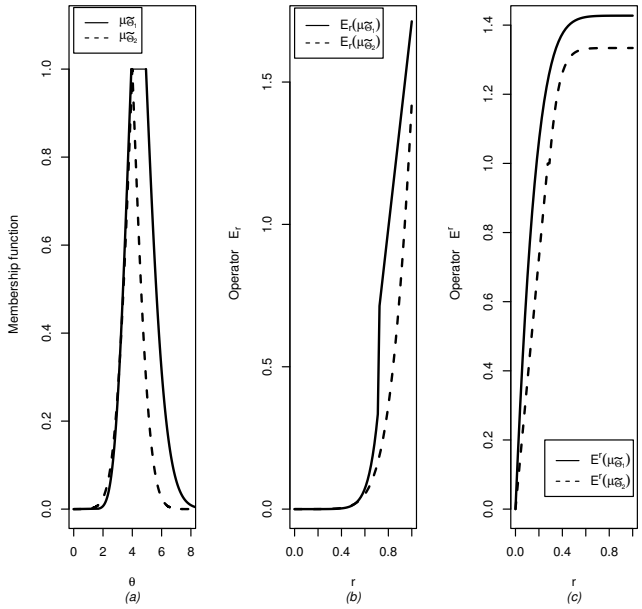


Figure 3. (**a**) membership functions for the Poisson rate based on the pivotal quantity (solid line) and based on the normal approximation (dotted line); (**b**) the r-down integral operator for the respective memberships; (**c**) the r-up integral operator for the respective memberships. $n = 5$ and $\bar{x} = 4$ are considered.

4.3. Normal Distribution

Now we discuss the situation where the parameter of interest is a vector. For this, consider, $X_1, \ldots, X_n$ to be a random sample from a normal distribution with mean θ_1 and variance θ_2. One approximate confidence set of significance level α for $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ can be defined by (see [53]):

$$C_\alpha(\mathbf{x}) = \left\{\boldsymbol{\theta} \in \Theta;\ U(\boldsymbol{\theta}; \mathbf{x}) < \chi^2_{2,1-\alpha}\right\} \quad \text{where} \quad U(\boldsymbol{\theta}; \mathbf{x}) = \frac{n}{\theta_2}(\bar{x} - \theta_1)^2 + \frac{n}{2\theta_2^2}(s^2 - \theta_2)^2,$$

where s^2 is the sample variance with denominator n. From Lemma 1, the associated membership function is given by

$$\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1 - \chi^2(U(\boldsymbol{\theta};\mathbf{x});2).$$

Here,

$$\Theta_r = \mu_{\widetilde{\Theta}}^{-1}([r,1]) = \{\boldsymbol{\theta} \in \Theta;\ \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) \geq r\} = \{\boldsymbol{\theta} \in \Theta;\ \chi^2(U(\boldsymbol{\theta};\mathbf{x});2) \leq 1 - r\}$$

and

$$\Theta_r^c = \{\boldsymbol{\theta} \in \Theta;\ \chi^2(U(\boldsymbol{\theta};\mathbf{x});2) > 1 - r\}.$$

Then, the r-down and r-up integral operators are given by

$$E_r(\mu_{\widetilde{\Theta}}) = \int_{\Theta_r^c} \mu_{\widetilde{\Theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \int_{-\infty}^{\infty}\int_{0}^{\infty} (1 - \chi^2(U(\boldsymbol{\theta};\mathbf{x});2)) I_{\Theta_r^c}(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

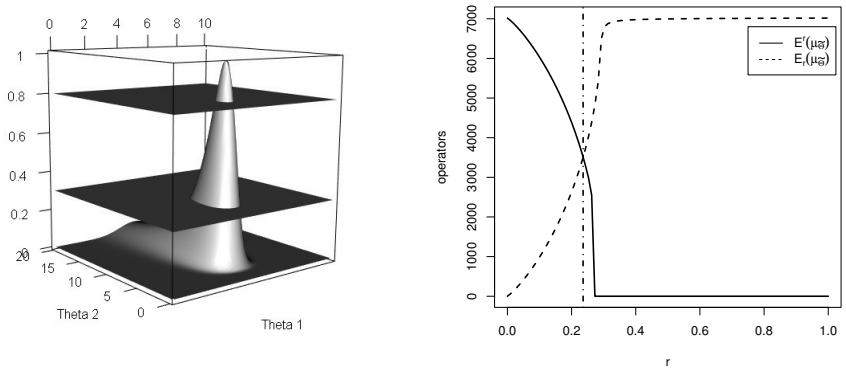
and

$$E^r(\mu_{\widetilde{\Theta}}) = \int_{\Theta_r} \mu_{\widetilde{\Theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \int_{-\infty}^{\infty}\int_{0}^{\infty} (1 - \chi^2(U(\boldsymbol{\theta};\mathbf{x});2)) I_{\Theta_r}(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where $I_A(\boldsymbol{\theta})$ is the indicator function.

These integrals can be computed numerically by using any software (for instance, `MAPLE`, `MATLAB`, `Ox`, `R`, `SAS`). Figure 4 presents the above membership function with $n = 5$, $\bar{x} = 5$, $s^2 = 5$ and the r-down and the r-up integral operators for the respective membership function.

In Figure 4a two planes are plotted, of height 0.3 and 0.8, to show the changes in confidence regions. In Figure 4b the graphs of (r, E_r) and (r, E^r) are plotted. It can be observed that $E_r = E^r$ when $r = 0.236$, as indicated by the dash-dotted vertical line.



(a) (b)

Figure 4. (**a**) membership function surface for mean and variance in a normal population; (**b**) the r-down and the r-up integral operators for the memberships. We consider $n = 5$, $\bar{x} = 5$ and $s^2 = 5$.

Basically, when the practitioner wants to compare general confidence sets, it is sufficient to plot the graphs of $(r, E_r(\mu_{\widetilde{\Theta}}))$ and $(r, E^r(\mu_{\widetilde{\Theta}}))$ in order to analyze the behaviors. Moreover, for univariate or bivariate confidence sets, the graph of $(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}))$ can be used in place of the usual confidence sets with a pre-fixed confidence level, since it brings much more information.

5. Applications

5.1. Confidence Set for Proportions in Bernoulli Trials

Following [54], we consider an experiment where N pairs of Bernoulli events denoted as A and B are observed. In this case, the outcomes are recorded as 1 (success) and 2 (failure), and the ith

observed pair is denoted by (Y_{i1}, Y_{i2}). The results of the experiment can be summarized in a 2×2 contingency table. Each n_{kl} corresponds to the number of pairs (Y_{i1}, Y_{i2}) with outcomes $Y_{i1} = k$ and $Y_{i2} = l$. Particularly, we consider a data set from [55] related to the study of the effect of the airway hyper-reactivity (AHR) before and after stem cell transplantation (SCT) in 21 children. The data are provided in Table 1.

Table 1. Airway hyper-responsiveness (AHR) status before and after stem cell transplantation (SCT) in 21 children.

(Event A) Before SCT/ (Event B) After SCT	AHR (Success)	No AHR (Failure)	Total
AHR (success)	$n_{11} = 1$	$n_{12} = 1$	2
No AHR (failure)	$n_{21} = 7$	$n_{22} = 12$	19
Total	8	13	$N = 21$

In this case, we are interested in a $(1 - \alpha)$ confidence interval for the difference of proportions $\theta = p_{12} - p_{21}$, where p_{12} is the proportion related to having AHR before SCT and not having it after SCT and p_{21} is the proportion related to the opposite event, namely, to not having AHR before SCT and having it after SCT. For constructing the confidence interval, we consider the following methods, namely, Wald, Wald with continuity correction, Wald with Agresti-Min pseudo-frequency adjustment and Wald with Bonett-Price Laplace adjustment (for further details about these methods, see [54,56,57]).

Denoting by $\widehat{\theta} = \widehat{p}_{12} - \widehat{p}_{21}$, where $\widehat{p}_{12} = n_{12}/N$ and $\widehat{p}_{21} = n_{21}/N$ are the ML estimates of p_{12} and p_{21} respectively and using the same notation as in Fagerland *et al.* [54], the confidence intervals for θ are as follows:

1. Wald's method
$$C_{1,\alpha}(\mathbf{x}) = \widehat{\theta} \pm \frac{z_{1-\alpha/2}}{N} \sqrt{n_{12} + n_{21} - (n_{12} - n_{21})^2/N}.$$
2. Wald's method with continuity correction
$$C_{2,\alpha}(\mathbf{x}) = \widehat{\theta} \pm \frac{z_{1-\alpha/2}}{N} \sqrt{n_{12} + n_{21} - (|n_{12} - n_{21}| - 1)^2/N}.$$
3. Wald's method with Agresti-Min pseudo-frequency adjustment
$$C_{3,\alpha}(\mathbf{x}) = \frac{\widetilde{n}_{12} - \widetilde{n}_{21}}{\widetilde{N}} \pm \frac{z_{1-\alpha/2}}{\widetilde{N}} \sqrt{\widetilde{n}_{12} + \widetilde{n}_{21} - (\widetilde{n}_{12} - \widetilde{n}_{21})^2/\widetilde{N}},$$
where $\widetilde{n}_{12} = n_{12} + 1/2$, $\widetilde{n}_{21} = n_{21} + 1/2$ and $\widetilde{N} = N + 2$.
4. Wald's method with Bonett-Price Laplace adjustment
$$C_{4,\alpha}(\mathbf{x}) = \widetilde{p}_{12} - \widetilde{p}_{21} \pm z_{1-\alpha/2} \sqrt{(\widetilde{p}_{12} + \widetilde{p}_{21} - (\widetilde{p}_{12} - \widetilde{p}_{21})^2)/(\widetilde{N} + 2)},$$
where $\widetilde{p}_{12} = (n_{12} + 1)/(N + 2)$ and $\widetilde{p}_{21} = (n_{21} + 1)/(N + 2)$.

For each method $\alpha \in I = [0, 1]$ and z_q is the qth-quantile of a standard normal distribution. Therefore, the membership functions for each method are given by:

1. Wald's method:
$$\mu_{\widetilde{\Theta}_1}(\theta) = 2\left(1 - \Phi\left(\frac{|\theta - \widehat{\theta}|N}{\sqrt{n_{12} + n_{21} - (n_{12} - n_{21})^2/N}}\right)\right)$$
2. Wald's method with continuity correction:
$$\mu_{\widetilde{\Theta}_2}(\theta) = 2\left(1 - \Phi\left(\frac{|\theta - \widehat{\theta}|N}{\sqrt{n_{12} + n_{21} - (|n_{12} - n_{21}| - 1)^2/N}}\right)\right)$$

3. Wald's method with Agresti-Min pseudo-frequency adjustment:

$$\mu_{\widetilde{\Theta}_3}(\theta) = 2\left(1 - \Phi\left(\frac{|\theta - (\widetilde{n}_{12} - \widetilde{n}_{21})/\widetilde{N}|\widetilde{N}}{\sqrt{\widetilde{n}_{12} + \widetilde{n}_{21} - (\widetilde{n}_{12} - \widetilde{n}_{21})^2/\widetilde{N}}}\right)\right)$$

4. Wald's method with Bonett-Price Laplace adjustment:

$$\mu_{\widetilde{\Theta}_4}(\theta) = 2\left(1 - \Phi\left(\frac{|\theta - (\widetilde{p}_{12} - \widetilde{p}_{21})|}{\sqrt{(\widetilde{p}_{12} + \widetilde{p}_{21} - (\widetilde{p}_{12} - \widetilde{p}_{21})^2)/(\widetilde{N} + 2)}}\right)\right)$$

Figure 5 shows the membership functions for each method presented before. In this case, we have that $\mu_{\widetilde{\Theta}_1} \ll_T \mu_{\widetilde{\Theta}_2}$. Therefore, by Theorem 3, we conclude that $\mathcal{C}_2 = \{C_{2,\alpha}(\mathbf{x})\}_{0\leq\alpha\leq1}$ is more conservative than $\mathcal{C}_1 = \{C_{1,\alpha}(\mathbf{x})\}_{0\leq\alpha\leq1}$. Consequently, Wald's method with continuity correction generates more conservative confidence intervals than Wald's method for all confidence levels. Similarly, $\mu_{\widetilde{\Theta}_4} \ll_T \mu_{\widetilde{\Theta}_3}$, that is, Wald's method with Agresti-Min pseudo-frequency adjustment generates more conservative confidence intervals than Wald's method with Bonett-Price Laplace adjustment for all confidence levels. Moreover, $\text{core}(\widetilde{\Theta}_1) = \text{core}(\widetilde{\Theta}_2)$ and $\text{core}(\widetilde{\Theta}_3) = \text{core}(\widetilde{\Theta}_4)$; however, $\text{core}(\widetilde{\Theta}_1) \neq \text{core}(\widetilde{\Theta}_3)$. This means that Wald's method with and without a continuity correction provides higher credibility to $\theta = -0.2857$ for the difference of proportions. However, Wald's method with Agresti-Min pseudo-frequency adjustment and with Bonett-Price Laplace adjustment gives a result greater than $\theta = -0.2608$.

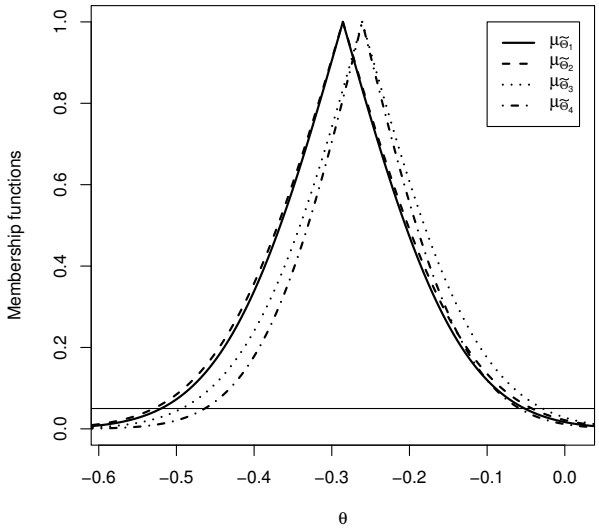


Figure 5. Membership functions for different confidence intervals.

The intersection of the continuous horizontal line with the membership functions shown in Figure 5 characterizes both, the lower and upper limits of a 95% confidence level according to each method described previously. Note that the largest differences are given in the lower limits.

Note also that, in this situation, the concept of total supremacy cannot be used to compare the membership functions of $\widetilde{\Theta}_1$ and $\widetilde{\Theta}_3$, since they have different cores. For this reason, we use Definition 9. The expressions for the $E_r(\mu_{\widetilde{\Theta}})$ and $E^r(\mu_{\widetilde{\Theta}})$ operators are similar to those given in Example 1, since their membership function shape is the same. Then,

$$E_r(\mu_{\widetilde{\Theta}}) = 4\sigma_*\left\{\frac{(c_{1r} - \widehat{\theta}_*)}{\sigma_*}\Phi\left(\frac{(c_{1r} - \widehat{\theta}_*)}{\sigma_*}\right) + \phi\left(\frac{(c_{1r} - \widehat{\theta}_*)}{\sigma_*}\right)\right\}$$

and

$$E^r(\mu_{\widetilde{\Theta}}) = 4\sigma_*\left\{\phi(0) - \frac{(c_{1r}-\widehat{\theta}_*)}{\sigma_*}\Phi\left(\frac{(c_{1r}-\widehat{\theta}_*)}{\sigma_*}\right) - \phi\left(\frac{(c_{1r}-\widehat{\theta}_*)}{\sigma_*}\right)\right\},$$

where $\widehat{\theta}_*$ and σ_* are as follows:

1. Wald's method
$$\widehat{\theta}_* = \widehat{\theta} \quad \text{and} \quad \sigma_* = \frac{1}{N}\sqrt{(n_{12}+n_{21}) - (n_{12}-n_{21})^2/N}.$$

2. Wald's method with continuity correction
$$\widehat{\theta}_* = \widehat{\theta} \quad \text{and} \quad \sigma_* = \frac{1}{N}\sqrt{(n_{12}+n_{21}) - (|n_{12}-n_{21}|-1)^2/N}.$$

3. Wald's method with Agresti-Min pseudo-frequency adjustment
$$\widehat{\theta}_* = \frac{\widetilde{n}_{12}-\widetilde{n}_{21}}{\widetilde{N}} \quad \text{and} \quad \sigma_* = \frac{1}{\widetilde{N}}\sqrt{(\widetilde{n}_{12}+\widetilde{n}_{21}) - (\widetilde{n}_{12}-\widetilde{n}_{21})^2/\widetilde{N}}.$$

4. Wald's method with Bonett-Price Laplace adjustment
$$\widehat{\theta}_* = \widetilde{p}_{12}-\widetilde{p}_{21} \quad \text{and} \quad \sigma_* = \frac{1}{\sqrt{\widetilde{N}+2}}\sqrt{\widetilde{p}_{12}+\widetilde{p}_{21} - (\widetilde{p}_{12}-\widetilde{p}_{21})^2}.$$

Figure 6 shows the $E_r(\mu_{\widetilde{\Theta}})$ and $E^r(\mu_{\widetilde{\Theta}})$ operators for Wald's method with and without Bonett-Price Laplace adjustment. Note that $\mu_{\widetilde{\Theta}_4}$ has r-down and r-up supremacy over $\mu_{\widetilde{\Theta}_1}$. Then, the interval $C_{1,\alpha}(\mathbf{x})$ is up/down-more conservative than $C_{4,\alpha}(\mathbf{x})$ for all confidence levels. Finally, Figure 7 depicts these operators for Wald's method with and without Agresti-Min pseudo-frequency adjustment. In this case, we have that $\mu_{\widetilde{\Theta}_3}$ has r-up and down-supremacy over $\mu_{\widetilde{\Theta}_1}$. Consequently, the interval $C_{1,\alpha}(\mathbf{x})$ is up/down-more conservative than $C_{3,\alpha}(\mathbf{x})$ for all confidence levels.

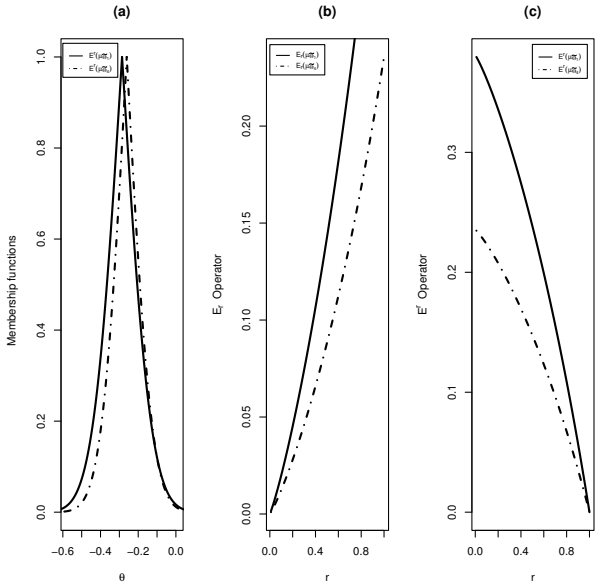


Figure 6. (**a**) Membership functions based on Wald's method, with and without Bonett-Price Laplace adjustment. (**b**) r-down integral operator for these memberships. (**c**) r-up integral operator for these memberships.

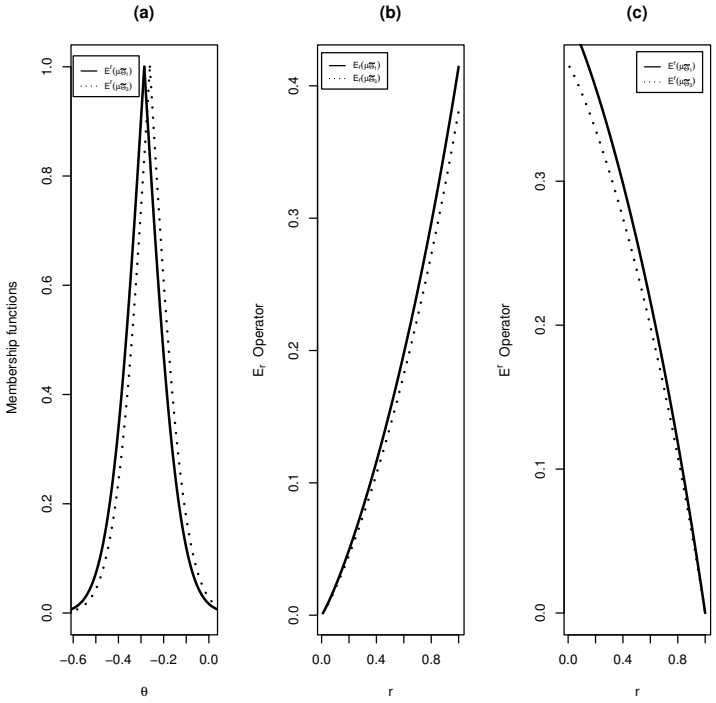


Figure 7. (**a**) Membership functions based on Wald's method with and without Agresti-Min pseudo-frequency adjustment. (**b**) *r*-down integral operator for these membership. (**c**) *r*-up integral operator for these membership.

5.2. Confidence Region for Regression Coefficients in a Normal Linear Model

In this section, we consider a dataset from [58] on features of Australian athletes available from the Australian Institute of Sport (AIS). This dataset has been analyzed previously by Arellano-Valle *et al.* [59], considering a linear regression model to study the relationship between lean body mass, height and weight of the Australian athletes. The model is given by

$$Lbm_i = \beta_1 Ht_i + \beta_2 Wt_i + \varepsilon_i, \quad i = 1, \ldots, 102,$$

where Lbm_i is the lean body mass, Ht_i is the height and Wt_i is the weight associated with $i = 1, \ldots, 102$ Australian male athletes. Table 2 presents a summary of the basic descriptive statistics for these variables.

Table 2. Descriptive statistics of the Australian athletes dataset: sample mean $\overline{x}$, sample standard deviation s, and sample skewness and kurtosis coefficients $\sqrt{b_1}$ and b_2, respectively.

Variable	$\overline{x}$	s	$\sqrt{b_1}$	b_2
Lbm	74.66	9.89	0.28	0.71
Ht	185.50	7.90	0.07	0.06
Wt	82.52	12.40	0.40	0.49

By assuming that the model error terms ϵ_i, $i = 1, \ldots, n$, are independent and normally distributed with constant variance σ^2 for each term, the following confidence region for β_1 and β_2 (see [60] for more details) is given by

$$\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{2\hat{\sigma}^2} \leq F(1 - \alpha, 2, 100),$$

where $\boldsymbol{\theta} = (\beta_1, \beta_2)^T$, $\hat{\boldsymbol{\theta}} = (\hat{\beta}_1, \hat{\beta}_2)^T$ is the maximum likelihood estimator for θ, $\hat{\sigma}^2$ is the maximum likelihood estimator for σ^2, $\mathbf{X}$ is the design matrix associated to the proposed regression model (assumed to be full column rank) and $F(q, p, s)$ is the qth-quantile corresponding to the Fisher-Snedecor's probability distribution with p and s degrees of freedom. The resulting membership function, based on the above confidence region, is

$$\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 1 - F\left(\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{2\hat{\sigma}^2}, 2, 100\right).$$

Figure 8 shows the membership function surface for the confidence ellipsoid for β_1 and β_2. This is the full information the confidence region can provide after observing this dataset.

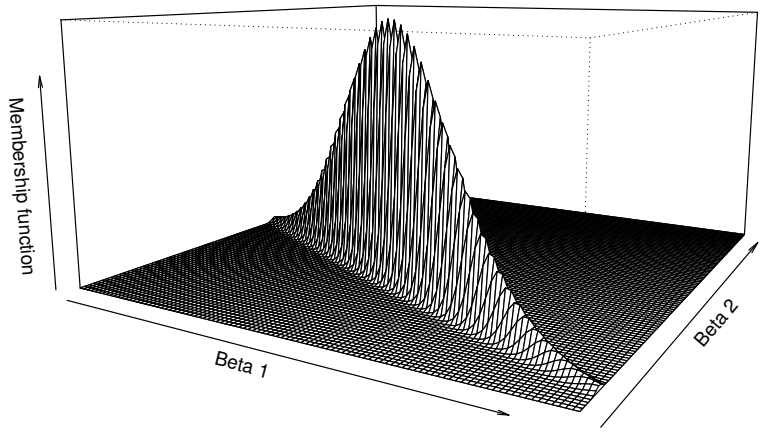


Figure 8. Membership function surface for β_1 and β_2 considering a linear regression model.

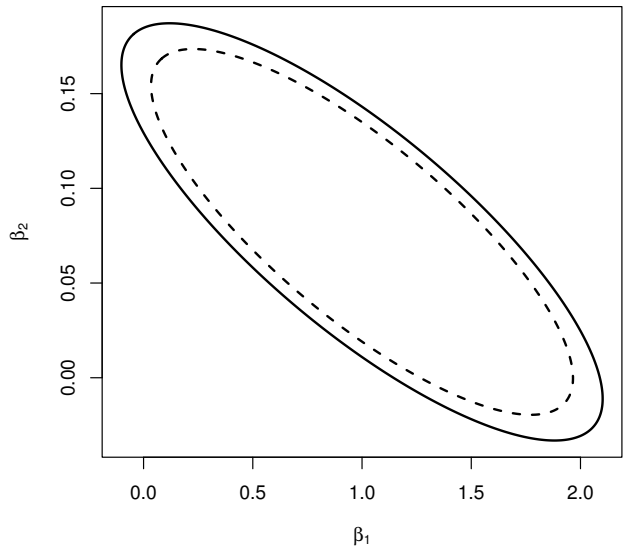


Figure 9. Confidence ellipsoid for β_1 and β_2 when $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 0.05$ (solid line) and $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 0.1$ (dashed line).

Notice that, by definition of the membership function above, $\mu_{\widetilde{\Theta}}(\hat{\boldsymbol{\theta}}) = 1$ and, by properties of the function $F(q, p, s)$, $\hat{\boldsymbol{\theta}}$ is the unique element of core$(\widetilde{\Theta})$. That is, $\hat{\boldsymbol{\theta}}$ is the unique value with full membership to the set $\widetilde{\Theta} = \{(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}));\ \boldsymbol{\theta} \in \Theta)\}$. Each $\boldsymbol{\theta}$ such that $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = k$ has membership of $k \times 100\%$. Figure 9 depicts the graphics $(\boldsymbol{\theta}, \mu_{\widetilde{\Theta}}(\boldsymbol{\theta}))$, where $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 0.05$ (solid line) and $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta}) = 0.1$ (dashed line), respectively; they are the contour curves of $\mu_{\widetilde{\Theta}}(\boldsymbol{\theta})$ at 0.05 and 0.1,

respectively. The values of $\boldsymbol{\theta}$ in solid line have membership of 5%, while the values of $\boldsymbol{\theta}$ in dashed line have membership 10%.

6. Discussion

The connection between confidence sets and fuzzy theory has already been investigated in the fuzzy literature. As mentioned, some works about this topic have been published in recent years (see for example among others [6–8,10]). All of them are related to the idea of generating a possibility distribution from the information provided by confidence sets. The main goal of this proposal is to find, in the context of fuzzy theory, more information in terms of degrees of possibility about the parameter of interest. More information can be extracted from confidence sets (or confidence intervals) after the sample is observed than just one fixed set (or interval). Indeed, a large family of sets is available.

There are also other proposals with the purpose of inferring about a parameter using confidence sets. These proposals are the fiducial inference [61], Dempster–Shaffer (DS) calculus [62], confidence distribution [46] and posterior Bayes. Some of them are carefully discussed by [63–65]. The main idea of fiducial inference is to consider a distribution for a parameter of interest using the sample information. The idea is to provide some probability statements about a parameter from the observed information, particularly, the information provided by a sufficient statistic. Fiducial distributions are often criticized because in some cases they do not integrate to one, *i.e.*, they are not proper probability distributions [66,67]. Moreover, some pivotal quantities used to build a fiducial distribution generate inconsistent results [68–70]. Therefore, some alternatives have been proposed such as confidence distributions and DS calculus.

The idea of confidence distributions is closely related with the fiducial distribution. As Schweder and Hjort [71] established, "confidence distributions are the Neymanian interpretation of Fisher's fiducial distributions" and they can be defined as a sample-dependent distribution able to represent confidence of all levels for a parameter of interest [46]. One of the most famous representatives of this class of distributions is the bootstrap distribution. It is important to remark that, the confidence distribution is considered a distribution estimator and its interpretation has to be done in a frequentist framework, considering a fixed and non-random parameter. Moreover, it is possible to obtain confidence intervals, point estimates and hypothesis test results about the parameter of interest using this distribution.

The DS calculus is based on the idea of convert observed data and pivotal relationships to upper and lower probability statements [72]. These statements are related to the probability of support by some subset of the parameter space, the contradiction of this event and the probability of "do not know" about both of them. According to Hanning and Xie [72], a confidence distribution can be formally put into a DS framework. The main difference between DS calculus and fiducial and confidence distributions is the concept of *degree of belief*. While the latter are focused on providing an estimator of a parameter of interest in terms of a quantity or/and interval/subset, DS calculus is concentrated on obtaining different degrees of belief or confidence for a simple question (related to a parameter). For this purpose, a *belief function* is used rather than probabilities.

In Bayesian inference we use the concept of prior distribution over the parametric space in order to infer about the parameter of interest. Although in this case, a distribution is also obtained, the notion of unknown parameter considered in the approaches described above is replaced by the notion of random parameter. Moreover, in this setup, the prior distribution summarizes the available information about this parameter. Some other authors have comprehensively discussed the methods discussed in this section. For a detailed discussion, we refer to [73–75].

Finally, our proposal intends to use the fuzzy set theory to represent a confidence set and, in this context, to provide more information to the statistical community about a specific confidence set. We believe, this approach represents both the uncertainty and imprecision existent about the parameter of interest by using the possibility theory obtained from a simple confidence set.

7. Conclusion and Final Remarks

In this paper, we revisited the connection between confidence sets and fuzzy sets through a membership function and therefore possibility distributions. Some elementary definitions and properties of fuzzy theory were revisited in order to assist readers not familiar with these non-standard tools. The connection indicates that fuzzy theory formalism can be utilized to interpret and give basis for confidence sets after observing the sample. Zadeh [76] argued that probability theory and fuzzy theory are complementary rather than competitive. As stressed in the paper, after observing the sample, any confidence sets have probability zero or one to contain the parameter value, so probabilistic post-data interpretations are not useful. We incorporated a new form of interpretation in which the core$(\widetilde{\Theta})$ is the set of all parameter values for which the related probability distributions best explain the observed data according to $\mathcal{C}$. In terms of a belief statement: the closer θ is to core$(\widetilde{\Theta})$, the more we believe that $P^{\boldsymbol{\theta}}$ that best model, according to $\mathcal{C}$ (the observed data). Moreover, the proposed membership function delivers much more information than a confidence set with a pre-fixed confidence level. Concepts of supremacy to compare different confidence sets and confidence procedures through membership functions were introduced and studied. Applications to usual examples were offered. This paper is part of large project on the foundation of classic statistics. Other works are being developed on hypothesis testing, evidence measures, coherence and so forth.

Acknowledgments: José A. González acknowledges funding support from Convenio de Desempeño en Formación de Profesores UPA 1203. Luis M. Castro acknowledges funding support by CONICYT-Chile through BASAL project CMM, Universidad de Chile and Grant 2012/19445-0 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). The research of Víctor H. Lachos was supported by Grant 2011/305054-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2011/17400-6 from FAPESP-Brazil. The research of Alexandre G. Patriota was supported partially by Grant 200115/2015-4/PDE from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2014/25595-0 from FAPESP-Brazil.

Author Contributions: José A. González proved the main results, analyzed the examples and data; Luis M. Castro wrote and corrected the main text, discussed the results, analyzed the examples and data; Victor H. Lachos corrected the main text and discussed the results; Alexandre G. Partriota conceived the main idea, proved some results, and wrote and corrected the main text. All of the authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Series A Contain. Pap. Math. Phys. Character* **1922**, *222*, 309–368.
2. Neyman, J. On the problem of confidence intervals. *Ann. Math. Stat.* **1935**, *6*, 111–116.
3. Neyman, J. Note on an article by Sir Ronald Fisher. *J. R. Stat. Soc. Ser. B.* **1956**, *18*, 288–294.
4. Zadeh, L. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1999**, *100*, 9–34.
5. Zadeh, L. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353.
6. Dubois, D.; Kerre, E.; Mesiar, R.; Prade, H. Possibility Theory, Probability and Fuzzy Sets. In *Fundamentals of Fuzzy Sets;* The Handbooks of Fuzzy Sets Series; Dubois, D., Prade, H., Eds.; Kluwer: Dordrecht, The Netherlands, 2000; pp. 385–386.
7. Dubois, D. Possibility theory and statistical reasoning. *Comput. Stat. Data Anal.* **2006**, *51*, 47–69.
8. Mauris, G. Possibility distribution: A unified representation for parameter estimation. In Proceedings of the Joint IFSA-EUSFLAT 2009 Conference, Lisbon, Portugal, 20–24 July 2009; pp. 1589–1594.
9. Dubois, D.; Prade, H. On Several Representations of an Uncertain Body of Evidence. In *Fuzzy Information and Decision Processes*; Gupta, M.M., Sanchez, E., Eds.; Elsevier: Amsterdam, The Netherlands, 1982; pp. 167–181.
10. Buckley, J. *Fuzzy Probability and Statistics*; Springer: Berlin/Heidelberg, Germany, 2006.
11. Patriota, A.G. A classical measure of evidence for general null hypotheses. *Fuzzy Sets Syst.* **2013**, *233*, 74–88.
12. De Barros, L.; Bassanezi, R. *Tópicos de Lógica Fuzzy E Biomatemática*; IMECC-Unicamp: Campinas, Brazil, 2006. (In Portuguese)

13. Coppi, R.; Gil, M.; Kiers, H. The fuzzy approach to statistical analysis. *Comput. Stat. Data Anal.* **2006**, *51*, 1–14.
14. Tanaka, H.; Vegima, S.; Asai, K. Linear Regression Analysis with Fuzzy Model. *IEEE Trans. Sustems Man Cybern.* **1982**, *2*, 903–907.
15. Wünsche, A.; Naether, W. Least-squares fuzzy regression with fuzzy random variables. *Fuzzy Sets Syst.* **2002**, *130*, 43–50.
16. Arabpour, A.; Tata, M. Estimating the parameters of a fuzzy linear regression model. *Iran. J. Fuzzy Syst.* **2008**, *5*, 1–19.
17. Geyer, C.J.; Meeden, G.D. Fuzzy and randomized confidence intervals and *p*-values. *Stat. Sci.* **2005**, *4*, 358–366.
18. Parchami, A.; Mashinchi, M.; Maleki, H.R. Fuzzy confidence interval for fuzzy process capability index. *J. Intell. Fuzzy Syst.* **2006**, *17*, 287–295.
19. Casals, M.; Gil, M.; Gil, P. The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information. *Eur. J. Oper. Res.* **1986**, *27*, 371–382.
20. Saade, J.J.; Schwarzlander, H. Fuzzy hypothesis testing with hybrid data. *Fuzzy Sets Syst.* **1990**, *35*, 197–212.
21. Saade, J.J. Extension of fuzzy hypothesis testing with hybrid data. *Fuzzy Sets Syst.* **1994**, *63*, 57–71.
22. Watanabe, N.; Imaizumi, T. A fuzzy statistical test of fuzzy hypotheses. *Fuzzy Sets Syst.* **1993**, *53*, 167–178.
23. Arnold, B.F. An approach to fuzzy hypothesis testing. *Metrika* **1996**, *44*, 119–126.
24. Arnold, B.F. Testing fuzzy hypotheses with crisp data. *Fuzzy Sets Syst.* **1998**, *94*, 323–333.
25. Taheri, S.; Behboodian, J. Neyman–Pearson Lemma for fuzzy hypotheses testing. *Metrika* **1999**, *49*, 3–17.
26. Filzmoser, P.; Viertl, R. Testing hypotheses with fuzzy data: The fuzzy *p*-value. *Metrika* **2004**, *59*, 21–29.
27. Dubois, D.; Prade, H. Operations on fuzzy numbers. *Int. J. Syst. Sci.* **1978**, *9*, 613–626.
28. Nasseri, S.; Taleshian, F.; Alizadeh, Z.; Vahidi, J. A New Method for Ordering LR Fuzzy Number. *J. Math. Comput. Sci.* **2012**, *4*, 283–294.
29. Zadeh, L. The concept of a linguistic variable and its application to approximate reasoning—I. *Inf. Sci.* **1975**, *8*, 199–249.
30. Sonka, M.; Hlavac, V.; Boyle, R. *Image Processing, Analysis, and Machine Vision*; Thomson: Toronto, Canada, 2008; Volume 3.
31. Wang, L. *A Course in Fuzzy Systems*; Prentice-Hall Press: Upper Saddle River, NJ, USA, 1999.
32. Goguen, J. L-fuzzy sets. *J. Math. Anal. Appl.* **1967**, *18*, 145–174.
33. Lee, K. *First Course on Fuzzy Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2005.
34. Zimmermann, H. *Fuzzy Set Theory and Its Applications*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 1992.
35. Dubois, D. *Fuzzy Sets and Systems: Theory and Applications*; Academic Press: Waltham, MA, USA, 1980.
36. Schervish, M. *Theory of Statistics*; Springer: New York, NY, USA, 1997.
37. Boes, D.; Graybill, F.; Mood, A. *Introduction to the Theory of Statistics*; McGraw Hill: New York, NY, USA, 1974.
38. Lehmann, E.; Romano, J. *Testing Statistical Hypotheses*; Springer: Berlin/Heidelberg, Germany, 2006.
39. Beran, R. Balanced simultaneous confidence sets. *J. Am. Stat. Assoc.* **1988**, *83*, 679–686.
40. Casella, G.; Berger, R. *Statistical Inference*; Duxbury Press: Belmont, CA, USA, 1990.
41. Cepeda-Cuervo, E.; Aguilar, W.; Cervantes, V.; Corrales, M.; Díaz, I.; Rodríguez, D. Intervalos de confianza e intervalos de credibilidad para una proporción. *Revista Colombiana Estadística* **2008**, *31*, 211–228. (In Spanish)
42. Fisher, R.A. Statistical methods and scientific induction. *J. R. Stat. Soc.* **1955**, *17*, 69–78.
43. Henderson, M.; Meyer, M.C. Exploring the confidence interval for a binomial parameter in a first course in statistical computing. *Am. Stat.* **2001**, *55*, 337–344.
44. Hoff, P. *A First Course in Bayesian Statistical Methods*; Springer: Berlin/Heidelberg, Germany, 2009.
45. Mauris, G.; Lasserre, V.; Foulloy, L. A fuzzy approach for the expression of uncertainty in measurement. *Measurement* **2001**, *29*, 165–177.
46. Xie, M.; Singh, K. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int. Stat. Rev.* **2013**, *81*, 3–39.
47. Fraïssé, R. *Theory of Relations*; Elsevier: Amsterdam, The Netherlands, 2000.
48. Owen, D.B. A table of normal integrals. *Commun. Stat. Simul. Comput.* **1980**, *9*, 389–419.

49. Ross, S. *Introduction to Probability and Statistics for Engineers and Scientists*; Academic Press: Waltham, MA, USA, 2009.
50. Guerriero, V. Power law distribution: Method of multi-scale inferential statistics. *J. Mod. Math. Front.* **2012**, *1*, 21–28.
51. Sahai, H.; Khurshid, A. Confidence intervals for the mean of a Poisson distribution: A review. *Biom. J.* **1993**, *35*, 857–867.
52. Correa, J.C. Interval Estimation for the Poisson Distribution Parameter with a Single Observation. *Revista Colombiana Estadística* **2007**, *30*, 69–75.
53. Arnold, B.; Shavelle, R. Joint confidence sets for the mean and variance of a normal distibution. *Am. Stat.* **1998**, *52*, 133–140.
54. Fagerland, M.; Lydersen, S.; Laake, P. Recommended tests and confidence intervals for paired binomial proportions. *Stat. Med.* **2014**, *33*, 2850–2875.
55. Bentur, L.; Lapidot, M.; Livnat, G.; Hakim, F.; Lidroneta-Katz, C.; Porat, I.; Vilozni, D.; Elhasid, R. Airway reactivity in children before and after stem cell transplantation. *Pediatr. Pulmonol.* **2009**, *44*, 845–850.
56. Newcombe, R. Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat. Med.* **1998**, *17*, 873–890.
57. Newcombe, R. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* **1998**, *17*, 857–872.
58. Cook, R.; Weisberg, S. *An Introduction to Regression Graphics*; Wiley: New York, NY, USA, 1994.
59. Arellano-Valle, R.; Castro, L.; Genton, M.; Gómez, H. Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Anal.* **2008**, *3*, 513–540.
60. Adkins, L.; Hill, R.C. An improved confidence ellipsoid for the linear regression model. *J. Stat. Comput. Simul.* **1990**, *36*, 9–18.
61. Fisher, R. Inverse probability. In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: Cambridge, UK, 1930; Volume 26, pp. 528–535.
62. Dempster, A. The Dempster-Shafer calculus for statisticians. *Int. J. Approx. Reason.* **2008**, *48*, 365–377.
63. Zadeh, L. Review of a mathematical theory of evidence. *AI Mag.* **1984**, *5*, doi:10.1609/aimag.v5i3.452.
64. Dubois, D.; Prade, H. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.* **1988**, *4*, 244–264.
65. Dubois, D.; Prade, H. Combination of Fuzzy Information in the Framework of Possibility Theory. In *Data Fusion in Robotics and Machine Intelligence*; Abidi, M., González, R., Eds.; Academic Press: San Diego, CA, USA, 1992; Chapter 12, pp. 481–505.
66. James, G. Discussion on the Symposium on Interval Estimation. *J. R. Stat. Soc. Ser. B* **1954**, *16*, 214–218.
67. Stein, C. An Example of Wide Discrepancy Between Fiducial and Confidence Intervals. *Ann. Math. Stat.* **1959**, *30*, 877–880.
68. Creasy, M. Limits for the Ratio of Means. *J. R. Stat. Soc. Ser. B* **1954**, *16*, 186–194.
69. Fieller, E. Some Problems in Interval Estimation. *J. R. Stat. Soc. Ser. B* **1954**, *16*, 175–185.
70. Mauldon, J. Pivotal Quantities for Wishart's and Related Distributions, and a Paradox in Fiducial Theory. *J. R. Stat. Soc. Ser. B* **1955**, *17*, 79–85.
71. Schweder, T.; Hjort, N. Confidence and likelihood. *Scand. J. Stat.* **2002**, *29*, 309–332.
72. Hanning, J.; Xie, M. A note on Dempster-Shafer recombination of confidence distributions. *Electron. J. Stat.* **2012**, *6*, 1943–1966.
73. Martin, R.; Zhang, J.; Liu, C. Dempster-Shafer theory and statistical inference with weak beliefs. *Stat. Sci.* **2010**, *25*, 72–87.
74. Martin, R.; Liu, C. Inferential models: A framework for prior-free posterior probabilistic inference. *J. Am. Stat. Assoc.* **2013**, *108*, 301–313.
75. Hannig, J. Generalized fiducial inference via discretization. *Stat. Sinica* **2013**, *23*, 489–514.
76. Zadeh, L. Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics* **1995**, *37*, 271–276.