Available online at www.sciencedirect.com

ScienceDirect

International Journal of Approximate Reasoning
47 (2008) 284–305

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Management of uncertainty in Statistical Reasoning: The case of Regression Analysis ☆

Renato Coppi *

Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma "La Sapienza", P.le A. Moro, 5 – 00185 Roma, Italy

Received 2 October 2006; received in revised form 26 February 2007; accepted 29 May 2007
Available online 4 July 2007

Abstract

Statistical Reasoning is affected by various sources of Uncertainty: randomness, imprecision, vagueness, partial ignorance, etc. Traditional statistical paradigms (such as Statistical Inference, Exploratory Data Analysis, Statistical Learning) are not capable to account for the complex action of Uncertainty in real life applications of Statistical Reasoning. A conceptual framework, called "Informational Paradigm", is introduced in order to analyze the role of Information and Uncertainty in these complex contexts. Regression Analysis is taken as the reference problem for developing the discussion. Three basic sources of Uncertainty are considered in this respect: (1) uncertainty about the relationship between response and explanatory variables; (2) uncertainty about the relationship between the observed data and the "universe" of possible data; (3) uncertainty about the observed values of the variables (imprecision, vagueness). Some of the available methods for coping with these different types of Uncertainty are discussed in an orderly way, from the simpler cases where only one source at a time is dealt with, to the more complex ones where all sources act together. Probabilistic and Fuzzy-Possibilistic tools are exploited, in this connection. In spite of the recent relevant contributions in this domain, the weaknesses and deficiencies of the current procedures for managing Uncertainty in Regression Analysis, as well as in other areas of Statistics, are emphasized. The elements of a generalized system of Statistical Reasoning, capable to deal with the various sources of Uncertainty, are finally introduced and the lines for future investigation in this perspective are indicated.

Keywords: Statistical Reasoning; Information processing; Uncertainty management; Informational Paradigm; Fuzzy analysis; Regression Analysis

1. Introduction

Statistical Reasoning is a specific, albeit relevant, instance of Approximate Reasoning or Reasoning under Partial Knowledge. Its distinctive features include: (1) applicability to "collective phenomena" (namely phe-

☆ This research was partially supported by grant "Prin 2005" of the Italian Ministry of University and Research (title: Models and methods to handle information and uncertainty in knowledge acquisition processes; coordinator: Professor G. Coletti). This contribution is gratefully acknowledged.

* Tel.: +39 0649910731; fax: +39 064959241.
E-mail address: renato.coppi@uniroma1.it

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2007.05.011

nomena which are characteristic of a *set* of observational instances, rather than of a single one); (2) reference to one or more "variables" as observed (or observable) on different "statistical units" (statistical data); (3) presence of several sources of Uncertainty, related to both the observational set up and the models utilized for analyzing the data; (4) uncertain conclusions of the reasoning process.

In the past century's literature, a few theoretical frameworks for developing Statistical Reasoning have been proposed. These range from the classical Inferential Paradigm, including the Bayesian approach (e.g., [1,2]) to the Descriptive-Exploratory Paradigm of the French school of thought usually referred to as "Analyse des Données" (e.g., [3]), to the more recent Statistical Learning Paradigm (e.g., [4,5]). However, all of the above frameworks do not allow for a complete treatment of the various sources of Uncertainty affecting the Statistical Reasoning process. In fact, the main source of Uncertainty possibly investigated in the mentioned contexts is "randomness" (quite often limited to the "data generation process" managed by means of appropriate probabilistic models). A broader view is taken by the Bayesian-subjectivistic approach which generalizes the use of probabilistic tools for managing uncertainties related to modeling assumptions (the so-called "prior information"). Then, the computation of posterior probability distributions, via the Bayes formula, allows the control of the uncertainty associated with the conclusions of the reasoning process. In spite of the increasing utilization of this approach in many applicative fields, still the need for a more inclusive treatment of Uncertainty in Statistical Reasoning is widely felt, by both theoretical statisticians and researchers in the various substantive domains. The main sources of Uncertainty that appear to be overlooked, in this connection, are "Imprecision" and "Vagueness". In a rather theoretical perspective the former notion of Uncertainty in the inferential process has been considered in the works of Dempster and Shafer, leading to the "Theory of Evidence" [6,7]. The use of "imprecise probabilities", in this respect, allows us to express the uncertainty associated with the selection of probability models or with the inferential conclusions of a statistical analysis (see also [8,9], and, in the specific field of Regression Analysis, [10]). Nevertheless, there remain unsolved problems concerning, for instance, the handling of Imprecision arising from the measurement tools utilized in analyzing real world phenomena. More generally, the Uncertainty stemming from the use of vague definitions or vague assessments in developing a statistical analysis should be carefully considered when evaluating the plausibility of the final inferences.

The basic contribution of Zadeh [11] introducing the notion of "Fuzzy Set" has opened the way to a new development of logical, mathematical and statistical thinking. In close connection with Probability Theory, Fuzzy Set Theory may provide the necessary tools for a generalized treatment of Uncertainty in Statistics, as we will try to argue in this paper.

Given a universe of reference, say U, a fuzzy set, $\tilde{A}$, is defined in U by means of its membership function:

$$\mu_{\tilde{A}}(u) \in [0,1] \quad \forall u \in U. \tag{1.1}$$

An equivalent way of characterizing $\tilde{A}$ is based on the notion of α-level sets. These are defined as follows:

$$[\tilde{A}]_\alpha = \{u \in U : \mu_{\tilde{A}}(u) \geqslant \alpha\} \quad \forall \alpha \in (0,1]. \tag{1.2}$$

On the basis of (1.2), the membership function of $\tilde{A}$ can be expressed in the following way:

$$\mu_{\tilde{A}}(u) = \sup_\alpha \alpha I\{[\tilde{A}]_\alpha\}(u), \tag{1.3}$$

where we denote by $I\{\cdot\}$ the usual characteristic function of a (crisp) set. The set

$$S = \{u \in U : \mu_{\tilde{A}}(u) > 0\}$$

is called the support of $\tilde{A}$. Usually, normalized fuzzy sets are utilized, verifying:

$$\exists u \in S : \mu_{\tilde{A}}(u) = 1.$$

When $U \equiv \mathbb{R}$ and we consider convex compact fuzzy sets $\tilde{X}$, i.e. satisfying:

$$[\tilde{X}]_\alpha = \text{closed and bounded interval} \quad \forall \alpha \in (0,1],$$

we get the important class of fuzzy intervals and, in particular, fuzzy numbers if $\mu_{\tilde{X}}(x) = 1$ only for a single point $x = x_0$ in the support of $\tilde{X}$.

A useful parametric family of fuzzy numbers is provided by the LR fuzzy numbers, $\tilde{X}$, whose membership function is given by

$$\mu_{\tilde{X}}(u) = \begin{cases} L\left(\frac{c-u}{l}\right), & u \leqslant c, \\ 1, & u = c, \\ R\left(\frac{u-c}{r}\right), & u \geqslant c, \end{cases}$$

where L and R are, respectively, the left and right "shape" functions, namely decreasing upper semicontinuous functions, satisfying: $L(0) = R(0) = 1$, $L(1) = R(1) = 0$, $L(z) = R(z) = 0 \ \forall z > 1$. The parameters c, l, r are respectively the center, left spread and right spread of $\tilde{X}$.

It is interesting to notice that the notion of *fuzziness* may be related to that of (graded) *possibility* (see [12, Chapters 5 and 8]). In fact, (1.1) can be interpreted as the "degree of possibility" of u, when we assume the viewpoint represented by the concept underlying the fuzzy set $\tilde{A}$.[1]

By means of this link, the logical and mathematical machinery of the theory of Fuzzy Sets may be usefully carried over into the framework of Possibility Theory (see, e.g., [13]), providing us with a powerful approach to cope with Uncertainty, using a non-additive measure (less restrictive than a probability measure).

In order to accommodate Fuzzy-Possibilistic and Probabilistic theories within the framework of Statistical Reasoning, we need a new Paradigm which could make this perspective conceptually meaningful. The previous paradigms do not seem to suit this purpose, due to their restricted conception of Uncertainty and Information. We propose the "Informational Paradigm" [14,15] as a more general epistemological framework for looking at Statistical Methodology in view of managing different types of Uncertainty affecting the processing of (statistical) Information.

Essentially, the Informational Paradigm looks at the Statistical Reasoning process as a logical system producing Information from Information. The elements entering this process are "informational ingredients", which can be distinguished in two categories: empirical and theoretical ingredients. The former ones, denoted by E, refer to (statistical) data collected from the real world in given observational or experimental contexts. The latter ones concern the various assumptions adopted in the reasoning process and the processed information constituting the output of the system ("Informational Gain"). These are indicated by the symbols A, A^p and P, according to whether they refer to basic (initial) assumptions, processing assumptions and processed information.

Each informational ingredient has a twofold nature. While it provides some kind of information, "covering" to some extent a lack of knowledge, at the same time it is affected by some sources of uncertainty (imprecision, vagueness, randomness, partial ignorance, etc.). Thus, Uncertainty and Information (in a Statistical Reasoning System) are inevitably linked with each other (Klir [12] expresses this in terms of "information-related uncertainty and uncertainty-related information"). In this connection, it is interesting to remark that three key contributions to this topic have been recently published by different Authors such as Bandemer [16], Klir [12] and Zadeh [17]. In particular, Klir and Zadeh make an important step toward the construction of a general conceptual, theoretical and technical frame for handling Uncertainty and Information in processes of Knowledge acquisition. The former Author utilizes the denomination "Generalized Information Theory", while the latter prefers to put the emphasis on the Uncertainty side, using the denomination "Generalized Theory of Uncertainty". Nonetheless, starting from different viewpoints, both tend to formalize a logical system capable to integrate Information and Uncertainty (see, e.g., the remarks in Chapter 10, note 10.7 of [12]).

Another noticeable line of thought aiming at unifying the treatment of Uncertainty and Information Processing under partial knowledge is witnessed by the works of Coletti and Scozzafava concerning the use of conditional probability in a coherent setting, in the spirit of De Finetti's subjectivistic approach (see, e.g., [18]). In this case conditional probability, looked on as a function of the conditioning "event", defines a general informational tool interpretable in Fuzzy-Possibilistic terms. It acts within an inferential system ruled by probability laws, whose intrinsic nature allows the assessment of uncertainty in probabilistic terms. However,

[1] For instance, if $\tilde{A}$ expresses the concept of "tall" (as referred to the height of a man), and $u = 1.80$ m, $(1.1) = 0.75$ may express the degree of possibility that a height of 1.80 m characterizes what we mean by a "tall man".

the capability of this theory to encapsulate the Statistical Reasoning process, in its various forms, is still to be verified.

Our approach differs from the above mentioned theories in that it uses the word ''Information'' in a rather intuitive and qualitative fashion. In fact, the main objective of this paper is to provide a logical systematization of the Statistical Reasoning process, whereby the ''Informational matter'' is elaborated in presence of several types of Uncertainty associated with the various employed ingredients and with the substantive problem at hand. This systematization should allow us to capture the weaknesses and deficiencies of the thus far available systems and, therefore, to suggest the lines of future research, in view of setting up an ''integrated system of Uncertainty Management in Statistical Reasoning''.

In order to accomplish the above task, we illustrate our point of view by referring to the general problem of Regression Analysis in Statistics. In this respect, we must specify a few points characterizing our treatment of Regression in the present context.

First, the discussion of Regression methods is kept within the domain of classical parametric regression with particular reference to parameter estimation and predictive use of the model. Second, it is organized according to an increasing level of complexity as to the management of the various sources of Uncertainty. Third, at each successive step of this illustration only some specific relevant methods are taken into account, in view of characterizing the way of dealing with the given stage of complexity. Of course, in the extremely rich literature concerning Regression Analysis, there exist numerous other methods of both parametric and non-parametric type, which will not be discussed here essentially because they would not add further substantial contributions to the line of thought underlying this paper. So, for instance, alternative non-parametric approaches to Regression are not dealt with, in spite of their methodological and practical importance in real life applications. Among them we mention: regression trees, neural networks, spline regression, techniques based on the use of wavelets, predictive regression methods based on ''boosting'' and ''bagging'' (see, for instance [5,19,20]).

The paper is organized in the following way. In Section 2, we characterize the Regression Problem in terms of Information and Uncertainty, according to the Informational Paradigm. Then, in Section 3, we define in detail the essential informational ingredients of the Statistical Reasoning process underlying Regression Analysis. The related Uncertainty is dealt with at the most elementary levels (namely, levels 0 and 1). In particular, at level 1, the linear regression model is assumed and the classical Least Squares estimation procedure is considered. It is underlined that, in this framework, there does not exist a proper system of Uncertainty Propagation. Section 4 is devoted to level 2 Regression Analysis. This is associated to the management of uncertainty stemming from the possible imprecision or vagueness of the response variable. This uncertainty is expressed by means of a suitable parametric family of fuzzy sets. The parameters of this family are appropriately regressed on the explanatory variables, and estimated on the basis of a Least Squares criterion applied to a distance between fuzzy sets. A ''side model'' exploiting fuzzy arithmetic relationships is then considered in order to complete the uncertainty analysis concerning the regression coefficients. In Section 5, level 3 Regression Analysis is discussed. This concerns the management of uncertainty due to the relationship between the observed data and the ''universe'' of possible data and to its influence on the estimation of the regression coefficients and on the predictive use of the regression model. The classical probabilistic approach is first taken into consideration, along with its own Uncertainty Propagation system. Then, the possibilistic perspective is investigated. This involves the use of fuzzy regression coefficients whose estimation is based on the principles of ''minimum fuzziness'' of the response variable and of ''possibilistic containment'' of the observed data. Section 6, devoted to level 4 Regression Analysis, deals with the more complex case where Uncertainty is jointly due to the imprecision/vagueness of the response variable and to its randomness (uncertainty caused by the data generation process). The notion of Fuzzy Random Variable is utilized in this context. A procedure, based on works [21–23] for estimating the regression coefficients, is described, along with the associated process of Uncertainty propagation. Some important limitations concerning the inferential system and the management of Uncertainty, at level 4 of Regression Analysis, are stressed. This is discussed in a broader perspective in Section 7. The conceptual and methodological elements of a generalized system of Statistical Reasoning, capable to manage the various sources of Uncertainty, are pointed out in view of stimulating specific investigations aimed at overcoming the deficiencies of the currently available methodology.

2. Information and uncertainty in the Regression Problem

The basic problem in Regression Analysis is the study of the relationship (say $\mathfrak{R}$) between a variable Y (''Response'') and a set of variables $X_1, \ldots, X_p$ (''Predictive'' or ''Explanatory'' variables).

In abstract terms, we shall denote this relationship by

$$\mathfrak{R}(Y; X_1, \ldots, X_p). \tag{2.1}$$

Notice that $\mathfrak{R}$ is a ''dependence'' relationship: Y is thought of as a variable whose behaviour *depends* on the behaviour of $X_1, \ldots, X_p$. In this sense, $\mathfrak{R}$ is an asymmetrical relationship.

The study of $\mathfrak{R}$ has two main objectives:

(i) determining the ''structure'' of $\mathfrak{R}$;
(ii) using $\mathfrak{R}$ for ''predicting'' Y, conditionally on knowing the ''evaluation'' of $(X_1, \ldots, X_p)$.

In Statistical Reasoning, an ''observational setting'' for the above problem is laid out. This consists in making available a set of ''evaluations'' of Y and $(X_1, \ldots, X_p)$ in n ''observational instances'':

$$[ev_i(Y), ev_i(X_1, \ldots, X_p)], \quad i = 1, \ldots, n,$$

where $ev_i(\cdot)$ denotes an evaluation of the variables in the argument, obtained on instance i. This evaluation consists of an assessment of the concerned variables, which may take the form of numerical values, linguistic expressions (such as ''good'', ''medium'', etc.), sets of numerical values (e.g., intervals on the real line), etc.

In general any given ''problem'', in scientific investigation as well as in decision making and also in everyday life, involves: (a) some kind of ''Information''; (b) some kind of associated ''Uncertainty''. The embedding of the problem in a specific ''Statistical Reasoning System'' adds more information and more uncertainties at the same time.

As mentioned in Section 1, with reference to the Informational Paradigm, the information getting through the system can be represented as a set of ''informational ingredients'' (whose nature may be empirical or theoretical). Each informational ingredient has a twofold link with uncertainty: on one hand it ''covers'' a part of uncertainty in the system (e.g., an observed datum partially covers ignorance about the physical world); on the other hand, however, it adds some kind of uncertainty in the system (e.g., the datum may be imprecise or vaguely defined). Thus, we may state that each informational ingredient *adds* uncertainty (related to itself), while *covering* some piece of uncertainty in the system.

The interplay between information and uncertainty in a Statistical Reasoning System yields two related propagation processes: (a) propagation of information (from initial information (E, A), to processed information (P)); (b) propagation of uncertainty (through the various uncertainty evaluations $ev[U(\cdot)]$ related to the different informational ingredients processed in the system).

A suitable Statistical Reasoning System should include an appropriate management of both information and uncertainty propagation processes.

In this connection, it should be noticed that the management of uncertainty involves in itself processing some kind of information (e.g., specific measures of uncertainty) to which there is inevitably associated some further uncertainty. Therefore, in order to avoid a paradox of ''regressio ad infinitum'', we convene to stop at some specified point the investigation of uncertainty, accepting a ''residual'' amount of ignored uncertainty in our reasoning process.

In the sequel we will specifically focus upon the management of uncertainty in the Regression Problem. Into the involved systems of information propagation will not be given a specific insight; they will be rather looked at as the necessary support for the propagation of uncertainty. We will examine the uncertainty management in the Regression Problem in a gradual manner, by distinguishing different successive ''levels'' of dealing with uncertainty in the regression framework, starting from basic uncertainties concerning the initial ingredients E and A and then deepening our insight into these ones along with managing the uncertainties linked with the processing assumptions (A^p) and other pieces of information fed into the system.

3. Management of uncertainty in Regression Analysis: levels 0 and 1

3.1. Level 0

Information

At level 0, the only Information we assume concerns the theoretical ingredient A. We denote this by A_{01}, where the subscript 0 refers to the "level" and 1 to the ordering of the various initial theoretical assumptions we are going to introduce:

A_{01}: we assume the "existence" of a "Relationship", $\mathfrak{R}$, between a phenomenon Y and p phenomena $X_1, \ldots, X_p$.

Notice that no empirical information is available at this stage ($E_{01} = \emptyset$).

Uncertainties

U_{01}: this uncertainty is related to $\mathfrak{R}$. In fact, at level 0, we do not formulate specific assumptions about the type and form of $\mathfrak{R}$.

U_{02}: this concerns the "observational instances" of $\mathfrak{R}$. Since E_{01} is empty, we have a complete uncertainty in this respect.

We may observe that both U_{01} and U_{02} represent a state of total ignorance. The only information at our disposal, at this starting level, concerns the logical existence of $\mathfrak{R}$.

3.2. Level 1

Level 1 is the initial stage of any procedure of statistical data analysis, namely the collection of data.

Information

Using the previously introduced notation for the informational ingredients and the corresponding uncertainties, we now consider the empirical information consisting of the statistical data and correspondingly update the set of uncertainties:

E_{11}: we get n "observational instances" concerning the evaluations made on $Y, X_1, \ldots, X_p$ in n instances (the data D):

$$D = \{[ev_i(Y), ev_i(X_1), \ldots, ev_i(X_p)], i = 1, \ldots, n\}.$$

Uncertainties

U_{11} ($\equiv U_{01}$): initial uncertainty w.r.t. $\mathfrak{R}$.

U_{12}: this is related to the relationship between D and the "universe" of "possible" instances (is D a sample from some population? What kind of sample is it? How is D generated? Etc.).

U_{13}: this uncertainty is associated with the evaluations made in D (e.g., these may be numerical and precise, quantitative but imprecise, linguistically expressed, vague).

Now, in order to cope with the above uncertainties while exploiting the empirical information E_{11}, we have to introduce a series of assumptions, including the "basic" ones (concerning the initial theoretical and empirical information, i.e. A_{01} and E_{11}) and some processing assumptions allowing us to manipulate the so far available informational material. The above mentioned series of assumptions constitutes a piece of further theoretical information (beside A_{01}), characterizing level 1 of analysis.

Theoretical information (at level 1)

Basic assumptions

A_{11} ($\equiv A_{01}$): logical existence of $\mathfrak{R}$.

A_{12}: the n observational instances in E_{11} are just "cases" of possible observations. No particular structure is assumed for the "universe" (see uncertainty U_{12}).

A_{13}: the data in D are assumed to be numerical and precise (i.e. we assume to deal with a crisp response and crisp explanatory variables) (see uncertainty U_{13}).

A_{14}: some kind of "form" is given to $\mathfrak{R}$ (see uncertainty U_{11}). A typical example, which we will take into consideration in the sequel, is the following *linear regression model*:

$$Y \cong g(\mathbf{x}, \boldsymbol{\vartheta}), \quad g(\cdot,\cdot) \in R, \tag{3.1}$$

where the class R of functions is given by

$$R = \{g(\mathbf{x}, \boldsymbol{\vartheta}) = [\mathbf{f}(\mathbf{x})]'\boldsymbol{\vartheta} : \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})]', \boldsymbol{\vartheta} \in \mathbb{R}^k\} \tag{3.2}$$

and $f_1(\mathbf{x}),\ldots,f_k(\mathbf{x})$ are given "design functions", namely appropriately chosen functions of the explanatory variables. Notice that the members of class R are uniquely identified by the vector of parameters $\boldsymbol{\vartheta}$.

By assuming A_{12} and A_{13} we are practically "nullifying" the corresponding uncertainties U_{12} and U_{13}. In fact, in the former case, we avoid introducing a model accounting for the relationship between observed and possible data (e.g., a data generation model). Of course, this will have relevant consequences on the possibility of managing the uncertainty propagation process, as we will see later. Likewise, in the latter case, we avoid doubting of the reliability of the measurements of variables Y and X_j's. However, this assumption may not stand several real life situations, where vagueness and imprecision inevitably affect our observational set up.

Differently from A_{12} and A_{13}, assumption A_{14} covers to some extent uncertainty U_{11}. But, at the same time, it introduces two further components of uncertainty: the first one refers to the approximation $\cong$ in (3.1); the other one concerns the specific values of vector $\boldsymbol{\vartheta}$.

Nonetheless, after having introduced some Processing Assumptions, we can see that U_{11} is effectively reduced, conditionally on the informational ingredients fed into the considered Statistical Reasoning System. We denote by a superscript "r" the reduced uncertainties. So, in our case, we have

$$U_{11}^r = (U_{11} \mid A_{11}, A_{12}, A_{13}, A_{14}, E_{11}, \text{ Processing Assumptions}) \tag{3.3}$$

as the residual uncertainty about $\mathfrak{R}$. We observe that U_{11}^r has two components: one is "internal" w.r.t. the set of assumptions; the other one is "external". The internal component is related to the uncertainties associated to the introduced assumptions (e.g., the "true" value of $\boldsymbol{\vartheta}$, or the measure of the approximation $\cong$, or the validity of A_{12} whose failure may produce a bias in the outputs of the system). The external component refers to the plausibility of A_{14}, which strongly restricts the universe of possible relationships in $\mathfrak{R}$.

Whileas a partial evaluation of the internal component is feasible within the given reasoning system, in order to assess (at least approximately) the external component we should widen the statistical system, covering, to some degree, the gap between A_{14} and $\mathfrak{R}$ (for instance, by considering a larger class of relationships including, for example, nonlinear regression models, neural networks, regression trees, etc.).

Processing assumptions

Under the above mentioned Basic Assumptions, and with particular reference to A_{14}, involving model (3.1), we introduce the following Processing Assumptions, allowing us to cope with the two components of (internal) uncertainty linked with A_{14} (namely those referring respectively to $\cong$ and $\boldsymbol{\vartheta}$):

A_{11}^p: we use a Least Squares (LS) "fitting criterion" which allows us to "quantify" the approximation in "$\cong$" and, at the same time, to produce an "estimate of $\boldsymbol{\vartheta}$:

$$\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{n} d^2[y_i, g(\mathbf{x}_i, \boldsymbol{\vartheta})]. \tag{3.4}$$

A_{12}^p: we adopt a specific "distance", d, to be plugged into (3.4).

Now, the informational set up for Regression Analysis at level 1 is completed. This consists of the following set of theoretical and empirical ingredients:

$$\mathfrak{I}_1 = (A_{11}, A_{12}, A_{13}, A_{14}, E_{11}; A_{11}^p, A_{12}^p). \tag{3.5}$$

Two kinds of Processed Information can be drawn from $\Im_1$. The first one, which we call P_{11}, refers to the estimate of ϑ given $\Im_1$ (usually denominated "LS estimate"). The second one, denoted by P_{12}, concerns the prediction of the unknown value y_h in a new observational instance, h, for which we know only the value taken by the explanatory vector: $\mathbf{x}_h$.

Processed information (at level 1)

P_{11}: given $\Im_1$, the inference on ϑ is expressed by the following "point estimate"

$$\hat{\vartheta} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}, \tag{3.6}$$

where

$$\mathbf{F} = \begin{bmatrix} f_1(\mathbf{x}_1) & \cdots & f_k(\mathbf{x}_1) \\ \vdots & & \vdots \\ f_1(\mathbf{x}_n) & \cdots & f_k(\mathbf{x}_n) \end{bmatrix}$$

is the design matrix.

P_{12}: given $\Im_1$ and $\mathbf{x}_h$, we get the following conditional estimate (predicted value) of y_h:

$$(\hat{y}_h|\Im_1, \mathbf{x}_h, P_{11}.) = \mathbf{f}_h'\hat{\vartheta}, \tag{3.7}$$

where $\mathbf{f}_h' = [f_1(\mathbf{x}_h), \ldots, f_k(\mathbf{x}_h)]$.

As mentioned before, both P_{11} and P_{12} are affected by uncertainty, as a result of the propagation of uncertainty through the Statistical Reasoning System of level 1. We denote these uncertainties, respectively, by

$$U(P_{11}|\Im_1) \tag{3.8}$$

and

$$U(P_{12}|\Im_1, \mathbf{x}_h, P_{11}). \tag{3.9}$$

It is clear that (3.8) and (3.9) are two sub-components of the internal component of the residual uncertainty U_{11}^r.

Given the informational ingredients of level 1 we are not able to make an evaluation of the uncertainties in (3.8) and (3.9). Even the Gauss–Markov theorem, concerning the BLUE (Best Linear Unbiased Estimator) property of the LS estimate of ϑ, cannot be used in this context, since assumption A_{12} excludes thinking in terms of sample and population and, consequently, of sampling distribution of $\hat{\vartheta}$.

Summing up, at level 1 of Regression Analysis, we are allowed to *reduce to some extent* the uncertainty about $\Re$, using P_{11}, and also the predictive uncertainty, using P_{12}. However, we cannot evaluate the uncertainty concerning these two elements of processed information. As a matter of fact, assumption A_{12} prevents us from using a data generation model which would enable us to manage the uncertainty propagation process by means of the classical mechanism: sampling model → sampling distribution of the statistics used for inferential purposes (e.g., estimators of the parameters) → measures of uncertainty based on these sampling distributions (e.g., the standard errors). On the other hand, there do not exist, within $\Im_1$, alternative possibilities for making other evaluations of (3.8) and (3.9).

We can conclude that, within the Statistical Reasoning System of level 1, there does not exist a system for managing the uncertainty propagation process. Under this point of view level 1 Regression Analysis constitutes an *incomplete* system of statistical reasoning, even if we restrict the scope of uncertainty management to the evaluation of uncertainty concerning processed information P_{11} and P_{12} (and thus overlooking many other important sources of uncertainty as outlined in this section).

4. Management of uncertainty in Regression Analysis: level 2

At level 2 of Regression Analysis we relax the basic assumption A_{13} of level 1, allowing for imprecision/vagueness in the measurement of the response variable Y (see uncertainty U_{13}). Although it would be possible

to relax also the assumption of crispness of the explanatory variables, we limit ourselves to arguing in terms of imprecisely/vaguely observed Y, without loosing in generality for our present purposes. Moreover, we keep the remaining basic assumptions of level 1, along with the empirical information E_{11} and the same basic uncertainties U_{11}, U_{12}, U_{13}.

Therefore, the initial informational set up of level 2 is as follows:

Empirical information

$E_{21} \equiv E_{11}$.

Theoretical information

$A_{21} \equiv A_{11} \equiv A_{01}$.

$A_{22} \equiv A_{12}$.

A_{23}: the data in D may be imprecise or vague. In particular we assume that Y is observed imprecisely or vaguely, whileas the X_j's are observed precisely (crisp variables). Furthermore, we assume that: $\forall i$, ev_i (Y) is a fuzzy variable defined on $\Re$, whose values belong to $\mathscr{F}_c^1$, namely the space of convex and compact fuzzy numbers (see [24]). Thus, our data become

$$\{(\tilde{y}_i, x_{1i}, \ldots, x_{pi}) : \tilde{y}_i \in \mathscr{F}_c^1,\ \mathbf{x}_i \in \mathbb{R}^p,\ i = 1, \ldots, n\}.$$

The above initial information is not sufficient for carrying out the analysis. To this purpose we have to introduce more ingredients (basic and processing assumptions). The literature on fuzzy Regression Analysis dealing with fuzzy response variables is rich of proposals. In this Section, we will restrict our considerations to an approach proposed by Coppi et al. [25], in a fuzzy Least Squares framework. We will only give an outline of the informational structure of this approach (informational ingredients and their treatment, management of the various sources of uncertainty). Consequently, we will not review the great variety of techniques that have been so far proposed, either in the LS or in the possibilistic perspectives, or even in a hybrid line of thought (see, for instance, [26]). The reader interested in getting a deeper insight into this methodological area (corresponding to level 2 of Regression Analysis) is addressed to the specific literature (e.g., see, for an overview, [27]).

4.1. Fuzzy least squares regression [25]

Assumption A_{23}, concerning the imprecision/vagueness of the response variable, is further specified in the following way:

A'_{23}: the fuzzy observations $\tilde{y}_i$ ($i = 1, \ldots, n$) are assumed to belong to the LR family of fuzzy numbers. Moreover, we look at them in a non-interactive way, considering the set of n membership functions:

$$\tilde{y}_i \equiv (y_i, l_i, r_i)_{LR}, \quad i = 1, \ldots, n, \tag{4.1}$$

where y_i, l_i, r_i are, respectively, the observed centers, left spreads and right spreads of the response variable, and LR denotes the "shape" of the specific LR membership function selected for our data. Therefore the *Empirical Information* E'_{21}, in this case, can be summarized by means of three vectors:

$\mathbf{y}' = (y_1, \ldots, y_n)$: centers-vector,

$\mathbf{l}' = (l_1, \ldots, l_n)$: left spreads-vector,

$\mathbf{r}' = (r_1, \ldots, r_n)$: right spreads-vector,

and by appropriate measures of the "shape" of the LR membership function. In the procedure we are illustrating (see [25]), these measures are given by

$$\lambda = \int_0^1 L^{-1}(\omega)\,\mathrm{d}\omega \quad \text{and} \quad \rho = \int_0^1 R^{-1}(\omega)\,\mathrm{d}\omega$$

(e.g., $\lambda = \rho = 1/2$ for triangular membership functions).

Consequently, in the analysis, the data are represented by

$$(\mathbf{y}, \mathbf{l}, \mathbf{r}, \lambda, \rho). \tag{4.2}$$

Notice that E'_{21} is inevitably affected by theoretical information concerning the type and shape of the membership function. However, it should be recognized that hardly we can think of a "purely" empirical information since any observation of the real world is necessarily conditional on some measurement device whose justification relies on theoretical assumptions.

In order to cope with uncertainty U_{11} (about $\Re$), we again make an assumption of type A_{14} by regressing the parameters of the response variable on the explanatory variables. This is formulated in the following Basic Assumption:

A_{24}: the relationship $\Re$ is expressed by means of regression equations involving the data represented in (4.2). Specifically, we assume:

$$\mathbf{y} \cong \mathbf{Fa} = \mathbf{y}^*, \tag{4.3}$$

$$\mathbf{l} \cong (\mathbf{Fa})b + \mathbf{1}_n d = \mathbf{l}^*, \tag{4.4}$$

$$\mathbf{r} \cong (\mathbf{Fa})g + \mathbf{1}_n h = \mathbf{r}^*, \tag{4.5}$$

where $\mathbf{a}$, b, g, d, h are unknown parameters, $\mathbf{F}$ is an appropriate design matrix of functions of the explanatory variables, and $\mathbf{y}^*$, $\mathbf{l}^*$, $\mathbf{r}^*$ denote the theoretical values of the centers, left and right spreads of the response variable in the n observations.

The uncertainty about $\Re$ (i.e. U_{11}) is now decomposed in two parts: (1) the uncertainty concerning model 4.3, 4.4 and 4.5, represented by the ignorance of the values taken by parameters $(\mathbf{a}, b, g, d, h)$ and by the approximation ($\cong$) of the model w.r.t. the data; (2) the residual uncertainty (not furtherly analyzed). The former component of uncertainty is managed by means of the Least Squares criterion, allowing us to minimize the approximation uncertainty while providing the estimates of the parameters. As usual, this requires the introduction of two processing assumptions, related, respectively, to an appropriate metric and a fitting criterion.

Processing assumptions

A^p_{21}: the distance between observed and theoretical (modeled) data is provided by the following Euclidean distance [28]:

$$\begin{aligned}\Delta^2_{LR}[(\mathbf{y},\mathbf{l},\mathbf{r})_{LR},(\mathbf{y}^*,\mathbf{l}^*,\mathbf{r}^*)_{LR}] &= \|\mathbf{y}-\mathbf{y}^*\|^2 + \|(\mathbf{y}-\lambda\mathbf{l})-(\mathbf{y}^*-\lambda\mathbf{l}^*)\|^2 + \|(\mathbf{y}+\rho\mathbf{r})-(\mathbf{y}^*+\rho\mathbf{r}^*)\|^2 \\ &= \Delta^2_{LR}[(\cdot),(\cdot)|\mathbf{a},b,g,d,h].\end{aligned} \tag{4.6}$$

A^p_{22}: the fitting criterion is the Least Squares criterion based on distance (4.6):

$$\min_{\mathbf{a},b,g,d,h} \Delta^2_{LR}[(\cdot),(\cdot)|\mathbf{a},b,g,d,h]. \tag{4.7}$$

The Informational set up at level 2, according to the present approach, is $\Im_2 = (A_{21}, A_{22}, A'_{23}, A_{24}, E'_{21}; A^p_{21}, A^p_{22})$.

Processed information

P_{21}: conditional on $\Im_2$, we get the iterative LS estimates $\hat{\mathbf{a}}, \hat{b}, \hat{g}, \hat{d}, \hat{h}$, which solve problem (4.7) (see [25]).

P_{22}: conditional on $\Im_2$, P_{21} and $\mathbf{x}_h$, we get a fuzzy estimate of $\tilde{y}_h$, given by

$$\hat{\tilde{y}}_h \equiv (\hat{y}_h, \hat{l}_h, \hat{r}_h)_{LR}, \tag{4.8}$$

where

$$\hat{y}_h = \mathbf{f}'_h\mathbf{a}, \tag{4.9}$$

$$\hat{l}_h = \hat{y}_h\hat{b} + \hat{d}, \tag{4.10}$$

$$\hat{r}_h = \hat{y}_h\hat{g} + \hat{h}, \tag{4.11}$$

in which $\mathbf{f}'_h = [f_1(\mathbf{x}_h), \ldots, f_k(\mathbf{x}_h)]$.

An Uncertainty Propagation process is feasible in the above framework, enabling us to evaluate the uncertainty pertaining to Processed Information P_{21} and P_{22}.

Uncertainty Evaluation (w.r.t. Processed Information)

$ev[U(P_{21}|\Im_2)]$: the following *implicit* fuzzy regression model underlies model (4.3)–(4.5):

$$\tilde{y}_i^{(*)} = \tilde{\beta}_1 \cdot f_1(\mathbf{x}_i) \oplus \cdots \oplus \tilde{\beta}_k \cdot f_k(\mathbf{x}_i), \quad i = 1, \ldots, n,$$

where $\oplus$ and $\cdot$ denote, respectively, the addition of fuzzy numbers and the multiplication of a fuzzy number by a scalar, and $\tilde{\beta}_j$ $(j = 1, \ldots, k)$ are fuzzy regression coefficients. If we assume that these parameters are expressed in terms of LR fuzzy numbers as follows:

$$\tilde{\beta}_j \equiv (\beta_j, \sigma_j, \tau_j)_{LR}, \quad j = 1, \ldots, k,$$

where β_j, σ_j and τ_j are, respectively, the center, left spread and right spread of the jth coefficient, then by using fuzzy arithmetic the following relationships are derived from the above *implicit* model:

$$\mathbf{y} = \mathbf{F}\boldsymbol{\beta}, \tag{4.12}$$

$$\mathbf{l} = |\mathbf{F}|\boldsymbol{\sigma}, \tag{4.13}$$

$$\mathbf{r} = |\mathbf{F}|\boldsymbol{\tau}, \tag{4.14}$$

where $|\mathbf{F}|$ denotes the matrix of absolute values $|f_{ij}|$. It is obvious that system (4.12)–(4.14) may not be verified by the estimates (4.9)–(4.11). However, we can use the relationships in (4.12)–(4.14) in order to get an estimate of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ which is "compatible" with the estimates (4.9)–(4.11). In fact, by assuming that the latter estimates "approximate" (up to a residual quantity) the fuzzy arithmetic relationships represented by system (4.12)–(4.14), in the following way:

$$\mathbf{y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_1, \tag{4.15}$$

$$\mathbf{l} = |\mathbf{F}|\boldsymbol{\sigma} + \boldsymbol{\varepsilon}_2, \tag{4.16}$$

$$\mathbf{r} = |\mathbf{F}|\boldsymbol{\tau} + \boldsymbol{\varepsilon}_3, \tag{4.17}$$

the above mentioned "compatibility" may be expressed in terms of Least Squares estimates of $\boldsymbol{\beta}$,$\boldsymbol{\sigma}$,$\boldsymbol{\tau}$ from system (4.15)–(4.17), using $\hat{\mathbf{y}}, \hat{\mathbf{l}}, \hat{\mathbf{r}}$ given by (4.9)–(4.11). Thus, we can easily see that $\hat{\boldsymbol{\beta}} = \hat{\mathbf{a}}$. Moreover, the estimates of the spreads $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ can be obtained by constrained LS satisfying the non-negativity constraints: $\hat{\boldsymbol{\sigma}} \geqslant \mathbf{0}, \hat{\boldsymbol{\tau}} \geqslant \mathbf{0}$, utilizing the NNLS algorithm [29]. The above illustrated procedure allows us to assess the uncertainty concerning the estimate $\hat{\mathbf{a}}$ (derived from (4.7)), by means of its estimated spreads $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\tau}}$. No similar uncertainty analysis is so far available for the remaining parameters b, d, g and h.

$ev[U(P_{22}|\Im_2, P_{21}, \mathbf{x}_h)]$: the assessment of the uncertainty related to the predictive use of the Regression Model is quite straightforward. In fact, it is implied in the fuzzy estimate (4.8), whose membership function evaluates the uncertainty about P_{22}:

$$\mu_{\hat{y}_h}(z) = \begin{cases} L\left(\frac{\hat{y}_h - z}{\hat{l}_h}\right), & z < \hat{y}_h, \\ 1, & z = \hat{y}_h, \\ R\left(\frac{z - \hat{y}_h}{\hat{r}_h}\right), & z > \hat{y}_h. \end{cases}$$

5. Management of uncertainty in Regression Analysis: level 3

At level 3 of Regression Analysis we try to manage Uncertainties U_{11} and U_{12}, namely about $\mathfrak{R}$ and about the link between the data (D) and the "Universe" of possible data. Instead, we overlook Uncertainty U_{13}, by assuming (as at level 1) that the data in D are precisely observed (data crispness). For achieving the above task, we will adopt two different viewpoints: the traditional inferential probabilistic approach on one hand and, on the other hand, the "possibilistic" perspective. While the former approach is embedded in Probability Theory, whereby Uncertainty is managed by means of probabilistic tools; the former one is derived from Possibility Theory (see, for instance, [13]) and utilizes Fuzzy-Possibilistic tools for coping with the relevant Uncertainties.

5.1. Approach 1 (classical probabilistic regression)

Basic assumptions

$A_{31} \equiv A_{11}$.

A_{32}: the link between the data in D and the Universe of potential data is formalized by means of a "stochastic data generation process", represented by a family of Probability Distributions over the Sample Space (i.e. the space of all possible samples which can be drawn from the Universe in the given observational set up). The explanatory variables $X_1, \ldots, X_p$ are assumed to be non-stochastic (i.e. we consider the analysis conditional on fixed design points). Instead, the response variable Y is assumed to be stochastic. The following statistical model is considered:

$$\{\mathcal{Y}^n, \wp\}, \tag{5.1}$$

where $\mathcal{Y}^n$ represents the sample space for $\mathbf{y}$, and $\wp$ the family of Probability Distributions on $\mathcal{Y}^n$.
Notice that, through $\wp$, we are able to measure (in probabilistic terms) the Uncertainty U_{12} specifically related to the "origin" of D. Consequently, we are also enabled to assess the Uncertainty stemming from the predictive use of Regression. This is, in fact, connected with the "Generalization Horizon" of the inferential results obtained by means of this Approach. The "Generalization Horizon" is, here, essentially based on the data generation process.

$A_{33} \equiv A_{13}$.

A_{34}: the relationship $\mathfrak{R}$ is expressed in terms of "Conditional Expectation" of the response variable. Namely:

$$E(Y|\mathbf{x}, p \in \wp) = g(\mathbf{x}), \quad g \in G, \tag{5.2}$$

where G is a suitable class of (Regression) Functions.
In a classical semi-parametric model we set

$$G \equiv R \equiv \{g(\mathbf{x}, \boldsymbol{\vartheta}) = [\mathbf{f}(\mathbf{x})]'\boldsymbol{\vartheta}; \boldsymbol{\vartheta} \in \mathbb{R}^k)], \tag{5.3}$$

as in level 1 analysis (see A_{14}).

Notice that, in the above setting, the dominant Assumption is A_{32}. This affects the way $\mathfrak{R}$ is formulated (in terms of expectation) and, consequently, the associated "Uncertainty Propagation System", which is based on classical Statistical Inference.

A widely used specification of Assumptions A_{32}, A_{34} is the following "model for semi-parametric regression analysis":

$$\begin{cases} \wp = \left\{p(\mathbf{y}|\boldsymbol{\vartheta}, \sigma, \ldots), \boldsymbol{\vartheta} \in \mathbb{R}^k, \sigma \in \mathbb{R}^+\right\}, \\ p(\mathbf{y}|\cdot) = \prod_i p(y_i|\cdot), \\ E(Y|\mathbf{x}, \boldsymbol{\vartheta}) = [\mathbf{f}(\mathbf{x})]'\boldsymbol{\vartheta}, \\ \boldsymbol{\Sigma}_y = \sigma^2 \mathbf{I}_n, \end{cases} \tag{5.4}$$

where $p \in \wp$ depends on the relevant $k+1$ parameters $(\boldsymbol{\vartheta}, \sigma)$ and on other (possibly infinite) parameters which are not of interest for inferential purposes. Moreover, the assumptions of stochastic independence among the various observations, along with that of same variance of the response variable are made.

In order to utilize model (5.4) for getting useful Information and evaluating the respective Uncertainties, we introduce the two usual processing assumptions, in the Least Squares perspective.

Processing assumptions

A^p_{31}: use of a Distance on $\mathcal{Y}^n$, e.g., the usual Euclidean distance.
A^p_{32}: adoption of a Fitting Criterion, e.g., the LS criterion.

The resulting Informational set up at level 3, according to approach 1 is

$$\Im_{3,1} = (A_{31}, A_{32}, A_{33}, A_{34}, E_{31} \equiv E_{11},\ \text{Model (5.4)};\ A^p_{31}, A^p_{32}). \tag{5.5}$$

Processed information

P_{31}: under $\Im_{3,1}$, we get the following LS estimates of the relevant parameters in model (5.4):

$$\hat{\boldsymbol{\vartheta}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}, \tag{5.6}$$

$$\hat{\sigma}^2 = (n-p)^{-1}\sum_i (y_i - \hat{y}_i)^2, \tag{5.7}$$

where $\hat{y}_i = \mathbf{f}_i'\hat{\boldsymbol{\vartheta}}, i = 1, \ldots, n,\ (\mathbf{f}_i' = [f_1(\mathbf{x}_i), \ldots, f_k(\mathbf{x}_i)])$.

P_{32}: conditional on $\Im_{3,1}$, P_{31} and $\mathbf{x}_h$, we get the following predicted value of the response variable in unit h:

$$(\hat{y}_h | \Im_{31}, P_{31}, \mathbf{x}_h) = \mathbf{f}_h'\hat{\boldsymbol{\vartheta}}. \tag{5.8}$$

Uncertainty evaluation

$ev[U(P_{31}|\Im_{3,1})]$: under $\Im_{3,1}$ it is not possible to make a specific assessment of this uncertainty. Nonetheless, the Gauss–Markov theorem guarantees that $\hat{\boldsymbol{\vartheta}}$ is the Best Linear Unbiased Estimator (BLUE) of $\boldsymbol{\vartheta}$, thus ensuring that $U(\hat{\boldsymbol{\vartheta}}|\Im_{3,1})$ is somehow minimized (in a specific class of estimators).
$ev[U(P_{32}|\Im_{3,1}, P_{31}, \mathbf{x}_h)]$: in this case we may use the estimate $\hat{\sigma}^2$ for assessing the uncertainty associated to the prediction of y_h.

Notice that more powerful evaluations of Uncertainties related to Processed Information can be obtained if we restrict $\wp$ to a Parametric Family, such as the Multinormal with Independent Homoscedastic Components (Regression Model under Normal Theory). In this case, the classical Processing Assumption consists in the use of the Maximum Likelihood Principle leading to ML estimators, Likelihood Ratio Tests, and Confidence Intervals. Thus, the above Uncertainties can be evaluated (in probabilistic terms) by means of standard errors, confidence intervals, significance levels, etc.

Therefore, by specifying Assumptions A_{32}, A_{34} according to the Normal theory, we obtain a more complete Uncertainty Propagation process, which propagates in probabilistic terms (following the classical inferential rules of the "repeated sampling" inferential system, see [1]), the sampling uncertainty related to the data generation process to the uncertainties concerning the Processed Information.

However, it should be underlined that, in the above framework (even under Normal Theory) the "Generalization Uncertainty" linked to the validity of the estimated relationship $\Re$ *beyond* the design points $\mathbf{X}$ is not considered.

As a matter of fact, the Uncertainty about $\Re$ affects immediately the Predictive Uncertainty in both the "conditional" perspective $[Y|(X_1, \ldots, X_p)]$ and the "unconditional one $(Y, X_1, \ldots, X_p)$. The classical probabilistic regression approach can cope with the conditional perspective of Uncertainty U_{11}, but not with the unconditional one. Dealing with the latter would require the introduction of a Family of Joint probability Distributions on the sampling space of the random vector $(y, \mathbf{x})$, consequently modifying the Basic Assumption A_{32} (see, for instance, [30]).

Approach 2 (possibilistic regression)

The pioneering work of Tanaka et al. [31] has introduced Possibilistic Thinking in the field of Regression Analysis. Since then, numerous contributions have been made in the possibilistic perspective in the domain of Regression (see, e.g., [32]) as well as in other statistical areas, such as Cluster Analysis, Principal Components, Multidimensional Scaling (see, for instance, [33]). Here, we will only give a hint to the basic ideas underlying the Possibilistic Approach, as applied to level 3 Regression Analysis. This will obviously be done following the lines of the Informational Paradigm, emphasizing the relevant Informational ingredients and the process of Uncertainty management in the Possibilistic system of Statistical Reasoning.

Basic assumptions
$A'_{31} \equiv A_{11}$.

A'_{32}: the data in D are through of as being realizations of a "Universe of Possibilities", which is not formalized. The link between D and this Universe is expressed through the (possibilistic) "Data Covering Principle" (see later).

$A'_{33} \equiv A_{13}$ (the data are supposed to be crisp).

A'_{34}: a "Possibilistic Relationship" between Y and $X_1, \ldots, X_p$ is assumed. This is represented by a class of regression functions with Fuzzy Parameters:

$$R' = \{g(\mathbf{x} \mid \tilde{\boldsymbol{\vartheta}}), \tilde{\boldsymbol{\vartheta}} \in \mathscr{F}_c^k\}. \tag{5.9}$$

A commonly used class for g is the linear one

$$g(\mathbf{x} \mid \tilde{\boldsymbol{\vartheta}}) = x_1 \cdot \tilde{\vartheta}_1 \oplus \cdots \oplus x_p \cdot \tilde{\vartheta}_p \tag{5.10}$$

where we set $k = p$. The Possibilistic Regression Model is

$$\tilde{y} = g(\mathbf{x} \mid \tilde{\vartheta}). \tag{5.11}$$

It must be underlined that the approximation $\cong$, typical of the regression models so far considered, is now incorporated *within the fuzziness* of regression model (5.11). Moreover, the (fuzzy) uncertainty about the Regression Coefficients (expressed by the membership functions pertaining to the $\tilde{\vartheta}_j$'s), *propagates* to the (fuzzy) uncertainty concerning the crisp response variable Y, via the Fuzziness Propagation System (including the extension principle, the fuzzy arithmetic operations, etc.; see [24]).

The information consisting of A'_{31}–A'_{34} and of $E'_{31} \equiv E_{11}$ (the crisp data in D) must be integrated by specific processing assumptions, in order to get "good" estimates of $\tilde{\vartheta}$ and to manage the Uncertainty Propagation process.

Processing assumptions

A'^{p}_{31}: the criterion of "Minimum Fuzziness" of the estimated response variable is adopted.
A'^{p}_{32}: the "Possibilistic Data Covering Principle" is applied, consisting in ensuring, at a given "Possibilistic Confidence Level", that all the observed responses are "contained" in the estimated ones (where the notion of fuzzy insiemistic containment is utilized).
A'^{p}_{33}: a given level of "Possibilistic Confidence" is chosen.
A'^{p}_{33}: a parametric form for the membership functions of the (unknown) fuzzy regression parameters $\tilde{\vartheta}_j$, $j = 1, \ldots, p$, is assumed. In particular, the LR fuzzy numbers are utilized in this connection:

$$\tilde{\vartheta}_j \equiv (\vartheta_j, \gamma_j, \delta_j)_{LR}, \quad j = 1, \ldots, p, \tag{5.12}$$

where ϑ_j, γ_j, δ_j denote, respectively, the centers, the left and right spreads.

Thus, the Informational set up for Possibilistic Regression Analysis at level 3 is provided by

$$\mathfrak{I}_{3,2} = (A'_{31}, A'_{32}, A'_{33}, A'_{34}, E'_{31}; A'^{p}_{31}, A'^{p}_{32}, A'^{p}_{33}, A'^{p}_{34}).$$

We briefly describe the *Procedure* for elaborating $\mathfrak{I}_{3,2}$.

1. By the Extension Principle, the (assumed) fuzziness of the regression parameters is "propagated" to the fuzziness of the estimated response $\tilde{y}$, as follows:

$$\tilde{y} \equiv (y^*, l, r)_{LR} \tag{5.13}$$

with

$$\begin{cases} y^* = \boldsymbol{\vartheta}'\mathbf{x}, \\ \mathbf{l} = \boldsymbol{\gamma}'|\mathbf{x}|, \\ \mathbf{r} = \boldsymbol{\delta}'|\mathbf{x}|. \end{cases} \tag{5.14}$$

Therefore, the h-level set of $\tilde{y}$ is given by

$$[\tilde{y}]_h = [\boldsymbol{\vartheta}'\mathbf{x} - L^{-1}(h)\gamma'|\mathbf{x}|, \boldsymbol{\vartheta}'\mathbf{x} + R^{-1}(h)\delta'|\mathbf{x}|]. \tag{5.15}$$

2. On the basis of (5.13), (5.14), the overall fuzziness of the (theoretical) response variable (for the entire collective of observations) is

$$F = \sum_{i=1}^{n} (\gamma + \delta)'|\mathbf{x}_i|. \tag{5.16}$$

3. The observed (crisp) responses, y_i, must be contained in the h-level set (5.15) of the corresponding theoretical (fuzzy) response $\tilde{y}_i$, giving the following constraints:

$$C_i \equiv \begin{cases} \{y_i \geqslant \boldsymbol{\vartheta}'\mathbf{x}_i - L^{-1}(h)\gamma'|\mathbf{x}_i|\}, \\ \{y_i \leqslant \boldsymbol{\vartheta}'\mathbf{x}_i + R^{-1}(h)\delta'|\mathbf{x}_i|\}, \end{cases} \quad i = 1, \ldots, n. \tag{5.17}$$

4. The following *Linear Programming problem* must then be solved:

$$\begin{cases} \min_{\boldsymbol{\vartheta},\boldsymbol{\gamma},\boldsymbol{\delta}} F, \\ \text{under constraints } C_i \ (i = 1, \ldots, n), \ \boldsymbol{\gamma} \geqslant \mathbf{0}, \ \boldsymbol{\delta} \geqslant \mathbf{0}. \end{cases} \tag{5.18}$$

Processed information

P'_{31}: the solution of problem (5.18) provides the estimates of the Fuzzy Parameters

$$\hat{\tilde{\vartheta}}_j \equiv (\hat{\vartheta}_j, \hat{\gamma}_j, \hat{\delta}_j)_{LR}, \quad j = 1, \ldots, p. \tag{5.19}$$

P'_{22}: under $\mathfrak{I}_{3,2}$, P'_{31} and $\mathbf{x}_h$, we get the following *fuzzy predicted value* of the response variable for unit h:

$$\hat{\tilde{y}}_h \equiv (\hat{\boldsymbol{\vartheta}}'\mathbf{x}_h, \hat{\boldsymbol{\gamma}}'|\mathbf{x}_h|, \hat{\boldsymbol{\delta}}'|\mathbf{x}_h|)_{LR}. \tag{5.20}$$

Uncertainty evaluation (w.r.t. P'_{21}, P'_{22})

$ev[U(P'_{31}|\mathfrak{I}_{3,2})]$: the uncertainty concerning the Regression Coefficients in model (5.10) is immediately expressed by means of the membership functions $\mu_{\hat{\tilde{\vartheta}}_j}(\cdot)$, $j = 1, \ldots, p$ pertaining to estimates (5.19).
$ev[U(P'_{32}|\mathfrak{I}_{3,2}, P'_{31}, \mathbf{x}_h)]$: also the uncertainty associated with the prediction of the response variable, having observed $\mathbf{x}_h$, is naturally represented by the membership function $\mu_{\hat{\tilde{y}}_h}(\cdot)$ of the fuzzy estimate (5.20).

More generally, if we look at the basic Uncertainties U_{11} (w.r.t. $\mathfrak{R}$) and U_{12} (w.r.t. the relationship between observed data D and Universe of possible data), we can make the following remarks.

(1) By means of P'_{31} we manage the parametric component of U_{11}, conditionally on model (5.10).
(2) $ev[U(P'_{31}|\mathfrak{I}_{3,2})]$ allows us to assess the uncertainty concerning the piece of Information P'_{31} which "covers" a part of U_{11}, according to remark (1).
(3) The residual part of Uncertainty U_{11}, stemming from the ignorance about the validity of model (5.10) as representative of relationship $\mathfrak{R}$, is, in some sense, "covered" by adopting the Possibilistic Data Covering Principle (at least at the possibilistic confidence level h). In fact, as long as the data from the real world fall under our observation, the fuzziness incorporated in model (5.10) allows us to account for the observed relationship between $Y$and $X_1, \ldots, X_p$ by suitably enlarging (if necessary) the spreads of the estimated parameters $\tilde{\vartheta}_j$ $(j = 1, \ldots, p)$ (see, for instance, the considerations made in [34] in this connection).
(4) As to uncertainty U_{12}, in the above illustrated possibilistic framework we do not formalize the link between the data in D and the universe of possible data by means of a "data generation model", as we do in the probabilistic approach. However, the possibilistic evaluations of the uncertainties related

to the estimated parameters $\hat{\tilde{\vartheta}}_j$ ($j = 1,\ldots,p$) and to the conditional predictors $\hat{\tilde{y}}_h$ can provide us with an appraisal of U_{12}. What we lack, in the possibilistic approach that we are examining, is the capability of evaluating the possible variation of the estimates $\hat{\tilde{\vartheta}}_j$ (and consequently of the predicted values based on them) due to the observation process, namely to the acquisition of new Empirical Information from the real world.

6. Management of uncertainty in Regression Analysis: level 4

At level 4 of Regression Analysis we try to manage simultaneously all of the three types of basic Uncertainties U_{11}, U_{12} and U_{13}. Therefore, as compared with level 3, we now allow for imprecise or vague data, taking into consideration, in particular, fuzzy data. In the sequel, we start the discussion of this case adopting the approach of Näther [23] (see, also, [21,22]).

Basic assumptions

$A_{41} \equiv A_{11}$.

A_{42}: the observational instances are generated by a stochastic mechanism.

A_{43}: $Y$is evaluated as a Fuzzy Random Variable, while $X_1,\ldots,X_p$ are observed crisply.

A_{44}: starting from the class R of regression functions considered at level 1 (see (3.2)), we *fuzzify* the elements of R through the "fuzzification" of $\boldsymbol{\vartheta}$. Therefore, we consider the class of linear regression functions with fuzzy parameters

$$\{R^* = [\mathbf{f}(\mathbf{x})]'\tilde{\boldsymbol{\vartheta}}, \tilde{\boldsymbol{\vartheta}} \in \mathscr{F}_c^k\}. \tag{6.1}$$

In order to make Assumptions A_{42}, A_{43}, A_{44} work together we need to formalize a stochastic mechanism having fuzzy outcomes. In this perspective, we may adopt the notion of Fuzzy Random Variable (FRV), as proposed in [35] (see also [36]). We should underline that there exist in the literature other approaches which could be considered in order to manage probabilistic uncertainty concerning fuzzy statistical data (see, for instance, [37,38]). However, it is not in the scope of the present paper the problem of illustrating and comparing the different ways of formalizing probability models for fuzzy statistical variables. Our aim, in the present context, consists rather in pointing out the necessity of combining probability and fuzziness for managing uncertainty in statistical analysis.

Definition 1 (*Fuzzy Random Variable* [35,36]). Let (Ω, A, P) be a Probability Space. Then

$$\tilde{Y}|\Omega \to \mathscr{F}_c^d$$

is a FRV on $\Re^d$ if for any $\alpha \in [0,1]$, the α-level set $[\tilde{Y}]_\alpha$ is a convex compact random set.

We also need to define at least the first two moments (mean and variance) of a FRV. The following definitions are usually adopted in our context.

Definition 2 (*Aumann's Expectation*). Using Aumann's integral, we define the Expectation of $\tilde{Y}$ as follows:

$$E(\tilde{Y}) \in \mathfrak{I}_c^d \text{ is such that } \forall \alpha \in [0,1] : [E(\tilde{Y})]_\alpha = E([\tilde{Y}]_\alpha),$$

where $E([\tilde{Y}]_\alpha) = \{E(\varphi) : \varphi(\omega) \in [\tilde{Y}]_\alpha(\omega) P - \text{a.e.}, \varphi \in L^1(\Omega, \mathrm{A}, P)\}$.

Definition 3 (*Variance*). We define the variance of a d-dimensional FRV, using the support function, in the following way:

$$\operatorname{var}(\tilde{Y}) = d \int_0^1 \int_{S^{d-1}} \operatorname{var} S_{\tilde{Y}}(u,\alpha) v(\mathrm{d}u)\,\mathrm{d}\alpha.$$

In particular, for d=1 and assuming that $\tilde{Y}$ is an LR symmetric fuzzy variable, namely

$$\tilde{Y} \equiv (y,s)_{L,} \tag{6.2}$$

where y and s are, respectively, the center and the spread, we have

$$E(\tilde{Y}) \equiv [E(y), E(s)]_L, \tag{6.3}$$

$$\mathrm{var}(\tilde{Y}) = \mathrm{var}\, y + L_2 \mathrm{var}\, s, \tag{6.4}$$

where $L_2 = \int_0^1 [L^{-1}(\alpha)]^2 \, \mathrm{d}\alpha$.

In the light of the previous definitions and of formula (6.3), and recalling the classical Linear Regression Model for crisp response, we can now set up the following "Componentwise Fuzzy Linear Regression Model":

$$A_{45}: \begin{cases} E(\tilde{y}_i|\mathbf{x}_i) \equiv (E(y_i), E(s_i))_L, \\ E(y_i) = [\mathbf{f}(\mathbf{x}_i)]'\vartheta, \quad \vartheta \in \mathbb{R}^k, \quad i = 1, \ldots, n, \\ E(s_i) = [\mathbf{h}(\mathbf{x}_i)]'\gamma, \quad \gamma \in \mathbb{R}^{k'}. \end{cases} \tag{6.5}$$

Model (6.5) is constituted by two different linear models, for the centers and the spreads, respectively. Notice that, for the Spread Model, the following constraints are imposed:

$$\begin{cases} [\mathbf{h}(\mathbf{x}_i)]'\gamma > 0 \quad \forall i, \\ \mathbf{x}_i \in C \subseteq \mathbb{R}^p \quad \forall i, \end{cases} \tag{6.6}$$

where C is the domain of interest for the Regression Model.

For the analysis of model (6.5)–(6.6) we introduce the following assumptions:

Processing assumptions

A_{41}^p: we focus on the class of Linear Unbiased Estimators of ϑ and γ.
A_{42}^p: we restrict the membership function of the response variable to be L-symmetric (see (6.2)).
A_{43}^p: we adopt the classical optimality principle of "Minimum Variance" in the class of Linear Unbiased Estimators, therefore aiming at getting the BLUE estimators of ϑ and γ.

Summing up, the Informational set up for Regression Analysis at level 4, in the above described framework is provided by

$$\mathfrak{I}_4 = (A_{41}, A_{42}, A_{43}, A_{44}, A_{45}, E_{41}; A_{41}^p, A_{42}^p, A_{43}^p),$$

where

$$E_{41}: \left.\begin{matrix} \tilde{\mathbf{y}}' = (\tilde{y}_1, \ldots, \tilde{y}_i, \ldots, \tilde{y}_n) \\ \mathbf{X} \end{matrix}\right\} \equiv D, \tag{6.7}$$

with $\tilde{y}_i \equiv (y_i, s_i)_L$, $i = 1, \ldots, n$.

Under $\mathfrak{I}_4$, we consider two Linear Estimators, respectively of ϑ and γ:

$$\begin{cases} \hat{\vartheta} = \Lambda \mathbf{y}, \\ \hat{\gamma} = \Gamma \mathbf{s}, \ [\mathbf{h}(\mathbf{x})]'\hat{\gamma} > 0, \quad \forall \mathbf{x} \in C. \end{cases}$$

Unbiasedness of $\hat{\vartheta}$ and $\hat{\gamma}$ implies

$$\Lambda\mathbf{F} = \mathbf{I}_k; \Gamma\mathbf{H} = \mathbf{I}_{k'}, \quad \text{where } \mathbf{H} = \begin{bmatrix} \mathbf{h}'(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}'(\mathbf{x}_n) \end{bmatrix}.$$

Processed information

P_{41}: under $\mathfrak{I}_4$, and adopting the classical LS criterion, the following componentwise BLUE estimates of the parameters of model (6.5)–(6.6) are obtained (see [23]):

$$\left\{\hat{\boldsymbol{\vartheta}}^* = (\mathbf{F}'\boldsymbol{\Sigma}_y^{-1}\mathbf{F})\mathbf{F}'\boldsymbol{\Sigma}_y^{-1}\mathbf{y}, \right. \tag{6.8}$$

$$\left\{\hat{\boldsymbol{\gamma}}^* = (\mathbf{H}'\boldsymbol{\Sigma}_s^{-1}\mathbf{H})\mathbf{H}'\boldsymbol{\Sigma}_s^{-1}\mathbf{s}, \right. \tag{6.9}$$

where $\boldsymbol{\Sigma}_y^{-1}$ and $\boldsymbol{\Sigma}_s^{-1}$ are, respectively, the inverses of the covariance matrices of the observed centers and spreads of the response variable.

P_{42}: under $\mathfrak{I}_4$, P_{41} and $\mathbf{x}_h$, the predicted value of the response variable of unit h (conditionally on having observed $\mathbf{x}_h$), is provided by

$$\hat{\tilde{y}}_h \equiv ([\mathbf{f}(\mathbf{x}_h)]'\hat{\boldsymbol{\vartheta}}^*, [\mathbf{h}(\mathbf{x}_h)]'\hat{\boldsymbol{\gamma}}^*)_L. \tag{6.10}$$

Uncertainty evaluation

$ev[U(P_{41}|\mathfrak{I}_4)]$: since we expressed the Fuzzy Regression Model in terms of two crisp regression models involving crisp parameters, we are not able to express the uncertainty about $\hat{\boldsymbol{\vartheta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$ in fuzzy terms. Furthermore, we have no information about the sampling distribution of $\hat{\boldsymbol{\vartheta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$ (since we have not specified the form of the Probability Distribution of $\tilde{Y}$) and then we cannot determine standard errors or confidence intervals. Therefore, our uncertainty evaluation is restricted to the guarantee that $\hat{\boldsymbol{\vartheta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$ have the BLUE property.

$ev[U(P_{42}|\mathfrak{I}_4, P_{41}, \mathbf{x}_h)]$: the uncertainty about the conditional prediction of $\tilde{y}_h$ is immediately expressed by the membership function of $\hat{\tilde{y}}_h$. However, there remains the stochastic uncertainty about the estimates $\hat{\boldsymbol{\vartheta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$, beside the uncertainty concerning the choice of the L-symmetric membership function for $\hat{\tilde{y}}_h$.

In a more general perspective, we can make the following remarks concerning the illustrated approach.

(1) Uncertainty U_{12} (about the relationship between data and Universe) is managed by the semi-parametric family $\wp(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \ldots)$ of Probability Distributions on σ_c^n. In this respect, we are in the same situation as the one described at level 3, Approach 1, of Regression Analysis (see Section 5.1).

(2) Uncertainty U_{13} (about the evaluation of the response variable Y) is managed by assuming the L-symmetric family of membership functions for the observed responses $\tilde{y}_i$'s.

(3) Uncertainty U_{11} (about $\mathfrak{R}$) is managed by means of the double regression model for the centers and spreads of the response variable (6.5), (6.6). Within this model, and with reference to the class of linear estimators, $\hat{\boldsymbol{\vartheta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$ "optimally" cover this Uncertainty. However, due to lack of information about sampling distributions, no further insight into U_{11} is possible from a probabilistic viewpoint. Nor it is possible from a fuzzy viewpoint due to the already underlined characteristics of model (6.5), (6.6).

(4) With reference to U_{12}, the "Generalization Power" of the present approach is limited to the observed points $\mathbf{x}_i \in C$ $(i = 1, \ldots, n)$. Enlarging the "Generalization Horizon" to the Universe of possible points in C, would require the consideration of a Joint Probability Distribution for response and explanatory variables (see [23]).

(5) Although the approach to level 4 analysis, so far illustrated, constitutes a step forward in view of the construction of a "complete" system of Statistical Reasoning in the field of Regression Analysis (i.e. a system allowing for the management of the main sources of Uncertainty, as described in this paper), the following weaknesses should be underlined:
   1. The restriction to L-symmetric membership functions;
   2. The restriction of the analysis to the Estimation problem;
   3. The splitting of the Regression Model into two separate Models. This may involve failing to achieve a "global" optimum, as well as other limitations in Uncertainty evaluation as previously pointed out;
   4. The adoption of a theoretically restrictive inferential approach (the BLUE theory), which may turn out to be inappropriate in a "General Uncertainty Management" framework;
   5. Incompleteness of the Uncertainty Propagation System. In fact, in the present approach Probability and Fuzziness are *juxtaposed* rather than *integrated*. This is witnessed by the lack of probability statements on most of the informational ingredients (except for the fuzzy response variable) and, conversely, by the lack of fuzzy statements on probabilistic ingredients.

7. Final considerations

In the previous Sections, after having embedded the Regression Problem in the framework of the Informational Paradigm, we have illustrated various procedures of Statistical Reasoning for coping with the different sources of Uncertainty affecting the implementation of Regression Analysis.

Arguing in a formalized, albeit schematic, way, we may summarize the above mentioned Informational process by means of an Informational Function with five arguments:

$$\Im(E, A, A^p, P; U), \tag{7.1}$$

where E, empirical data; A, basic assumptions; A^p, processing assumptions; P, processed information (output of the reasoning process); U, uncertainty.

Notice that U acts on each of the four informational ingredients (or their combinations) and on additional theoretical ingredients pertaining to the specific problem at hand (in the case of Regression: relationship between Data and Universe, relationship between response and explanatory variables, etc.). Thus, we can consider the following Uncertainty Functions: $U(E)$, $U(A)$, $U(A^p)$, $U(P)$, $U[f(E, A, A^p, P)]$, U (theoretical ingredients of the problem).

Moreover, U may arise from different sources: randomness, imprecision, vagueness, partial or total ignorance, granularity of concepts, etc.. Basically, so far, three different theoretical approaches to coping with the above types of uncertainty have been proposed in the literature:

(1) Probability theory and the associated classical inferential techniques (either in the ''repeated sampling'' or in the ''likelihood principles'' frameworks); (2) Fuzzy-Possibilistic theory and the relative procedures for statistical analysis (see, e.g., [39]); (3) Interval Analysis as applied to Statistical Reasoning (see, for instance, [8]).

In the present paper, we focused, in particular, on approaches 1 and 2. In this respect, our (not exhaustive) analysis has emphasized that, up to now, the main lines of research in the field of Regression Analysis have either selected one specific approach to the management of uncertainty or, when considering both approaches 1 and 2, have proceeded by ''juxtaposing'' them. Namely, different roles have been assigned to the probabilistic and fuzzy ingredients of the analysis. In fact, in Section 6, we have shown some regression procedures where the empirical data are fuzzy but the models for analyzing them have a stochastic nature.

However, even in the specific perspective of ''juxtaposing'' the approaches to the management of uncertainty in a given Statistical Reasoning System, we can point out several topics needing a deeper insight. The following is just a first list.

(1) Definition and construction of Parametric Families of Fuzzy Random Variables (e.g., families playing the role of the Gaussian family, or, more generally, of the exponential family in the classical crisp context; it should be underlined, in this connection, that some Authors have suggested notions of ''Normal'' FRV's (e.g., [40]), which unfortunately, do not appear to have produced fruitful results in terms of inferential techniques based on them).
(2) Setting up ''complete'' inferential procedures for FRV's, including point and interval estimation, testing procedures and so on. As a matter of fact, at least in the Regression framework, the available techniques are restricted to point estimation (without assessment of the respective uncertainty) and to testing hypotheses in some simple cases. Of course, in this respect, we should look not only for ''exact'' or ''asymptotic'' theory and techniques, but also for more pragmatic approaches as those based on resampling (for a systematic approach utilizing ''bootstrap'' techniques for drawing inference on FRV's see [41–43]).
(3) More extensive use of the fuzzification of theoretical ingredients. For instance, with reference to $U(A^p)$ in Regression Analysis, we may consider the following extensions of the domain of Fuzzy Theory in Regression Analysis: (a) fuzzification of the ''entries'' of the different explanatory variables in the Regression Model (namely, we may think of a ''membership degree'' of an explanatory variable in the Model); (b) fuzzification of the ''entries'' of nonlinear effects of the explanatory variables (e.g., products), with the same criterion as in (a). Furthermore, in a theoretical perspective, the following aspects connected with the inferential use of the notion of FRV deserve a particular attention:

(4) Definition of key-notions of the probabilistic inferential reasoning such as: sufficiency, likelihood, etc.
(5) Development of a theory allowing the definition of "optimal" inferential procedures which could match the well known classical theories of efficient estimators, uniformly most powerful tests, and so on.
(6) As a consequence of (5), construction of "optimal" inferential procedures, based on statistical models for FRV's.

So far, we considered the "juxtaposition" approach, with particular reference to the probabilistic analysis of fuzzy data. However, a more general development can be envisaged, leading to a closer interaction among the various ways of dealing with uncertainty in Statistical Reasoning, and more specifically between Fuzziness and Randomness. In order to illustrate this point, let us briefly recall the classical probabilistic (parametric) model for statistical inference:

$$\{S^n, p(\mathbf{x} \mid \boldsymbol{\vartheta}), \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}, \tag{7.2}$$

where $\mathbf{x} \in S^n =$ sample space; $p(\cdot|\cdot) =$ family of probability distributions on S^n (sampling model); $\boldsymbol{\vartheta} \in \boldsymbol{\Theta} =$ parameter space.

The information contained in $\mathbf{x}$, concerning the unknown parameter vector $\boldsymbol{\vartheta}$, is expressed by the Likelihood Function:

$$l(\boldsymbol{\vartheta}|x) \propto p(\mathbf{x}|\boldsymbol{\vartheta}). \tag{7.3}$$

Often, like in the case of Regression Analysis, some or all of the parameters in $\boldsymbol{\vartheta}$ are reparametrized in terms of new parameters, say $\boldsymbol{\varphi} \in \boldsymbol{\Phi}$, expressing more directly the information we want to capture (e.g., the regression coefficients, given a set of explanatory variables). This may be represented by a "side statistical model": $\mathrm{M}(\mathbf{x}|\boldsymbol{\varphi})$. The likelihood for $\boldsymbol{\varphi}$ can be immediately derived from (7.3).

Now, using the notion of FRV's the following extension to the fuzzy setting could be envisaged:

$$\{\mathscr{F}^n, p(\mathbf{x}|\vartheta_1, \gamma_1), \vartheta_1 \in \Theta_1, \gamma_1 \in \Gamma_1\}, \tag{7.4}$$

$$\{M_F(\mathbf{x}|\vartheta_1, \gamma_1, \vartheta_2, \gamma_2), \vartheta_2 \in \Theta_2, \gamma_2 \in \Gamma_2\}, \tag{7.5}$$

where

σ^n	sample space of fuzzy values
$\mathbf{x}$	fuzzy observed data
$\boldsymbol{\vartheta}_1$	crisp parameters of sampling model
γ_1	fuzzy parameters of sampling model
M_F	fuzzy model for $\mathbf{x}$ (e.g., fuzzy regression model, fuzzy clustering model (see, e.g., [44] in a non-probabilistic setting), fuzzy latent structure model (see, e.g., [15] in a non-probabilistic setting)
$\boldsymbol{\vartheta}_2$	crisp parameters of M_F (e.g., crisp regression coefficients, crisp cluster prototypes, crisp factor loadings)
γ_2	fuzzy parameters of M_F (e.g., fuzzy regression coefficients, fuzzy cluster prototypes, fuzzy factor scores).

Further generalizations of model (7.4)–(7.5) might be considered. For instance, imprecise probabilities for $p(\cdot|\cdot)$ could be adopted (see, e.g., [9] for a treatment of imprecise probabilities). This would allow us to manage a specific source of uncertainty due to ignorance about the form of the probability model for our data generating process. Of course, it is quite a difficult task to give sound inferential foundations to statistical inference in this setting. One basic theoretical problem concerns the following question: is there a sensible counterpart of the notion of "likelihood function" in this framework? Should we consider a sort of "confidence function" embodying both the random and fuzzy uncertainties?

Before giving a satisfactory answer to the above question and to the more general foundational problem concerning the construction of an "integrated system of uncertainty management in Statistical Reasoning", some partial steps in this direction can be taken. An example is provided by some proposals concerning the fuzzification of inferential statements, such as fuzzy confidence intervals, fuzzy significance levels, fuzzy statistical decisions (see, e.g., [33]). An important theoretical and methodological feature of the above

mentioned integrated approach to uncertainty management is constituted by the "Uncertainty Propagation Systems". With reference to model (7.4)–(7.5), these should allow us to consistently propagate the various forms of uncertainty from the original informational inputs through the final inferential conclusions. Some pioneering attempts in this direction have been made (e.g., [45–47]). There is still much room for wide range investigations in view of establishing a "Generalized Management of Uncertainty in Statistical Reasoning", capable to integrate and expand the results and acquisitions of Probability Theory, Fuzzy-Possibilistic Thinking and Interval Analysis.

References

[1] D.R. Cox, D. Hinkley, Theoretical Statistics, Chapman and Hall, London, 1974.
[2] G. Casella, R.L. Berger, Statistical Inference, Duxbury Advanced Series, Pacific Grove, 2002.
[3] J.P. Benzécri, L'analyse des données, Dunod, Parigi, 1973.
[4] V. Vapnik, The Nature of Statistical Learning, Springer-Verlag, New York, 1996.
[5] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning – Data Mining Inference and Prediction, Springer-Verlag, New York, 2001.
[6] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.
[7] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[8] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[9] P. Walley, Towards a unified theory of imprecise probability, International Journal of Approximate Reasoning 24 (2000) 125–148.
[10] S. Petit-Renaud, T. Denoeux, Nonparametric regression analysis of uncertain and imprecise data using belief functions, International Journal of Approximate Reasoning 35 (2004) 1–28.
[11] L.A. Zadeh, Fuzzy sets, Information Control 8 (1965) 338–353.
[12] G. Klir, Uncertainty and Information – Foundations of Generalized Information Theory, John Wiley and Sons, New Jersey, 2006.
[13] D. Dubois, H. Prade, Possibility Theory, Plenum Press, New York, 1988.
[14] R. Coppi, A theoretical framework for data mining: the "Informational Paradigm", Computational Statistics & Data Analysis 38 (2002) 501–515.
[15] R. Coppi, P. Giordani, P. D'Urso, Component models for fuzzy data, Psychometrika 71 (2006) 733–761.
[16] H. Bandemer, Mathematics of Uncertainty – Ideas Methods Application Problems, Springer-Verlag, Berlin, Heidelberg, 2006.
[17] L.A. Zadeh, Generalized theory of uncertainty (GTU) – principal concepts and ideas, Computational Statistics & Data Analysis 51 (2006) 15–46.
[18] G. Coletti, R. Scozzafava, Probabilistic Logic in a Coherent Setting, Kluwer Academic Publishers, The Netherlands, 2002.
[19] L. Breiman, Bagging predictors, Machine Learning 24 (1996) 123–140.
[20] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.
[21] A. Wünsche, W. Näther, Least-squares fuzzy regression with fuzzy random variables, Fuzzy Sets and Systems 130 (2002) 43–50.
[22] W. Näther, On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data, Metrika 51 (2000) 201–221.
[23] W. Näther, Regression with fuzzy random data, Computational Statistics & Data Analysis 51 (2006) 235–252.
[24] H.J. Zimmermann, Fuzzy Set Theory and its Applications, Kluwer Academic Press, Dordrecht, 2001.
[25] R. Coppi, P. D'Urso, P. Giordani, A. Santoro, Least squares estimation of a linear regression model with LR fuzzy response, Computational Statistics & Data Analysis 51 (2006) 267–286.
[26] Y.H. Chang, B.M. Ayyub, Fuzzy regression methods – a comparative assessment, Fuzzy Sets and Systems 119 (2001) 187–203.
[27] P. Diamond, H. Tanaka, Fuzzy regression analysis, in: R. Slowinski (Ed.), Fuzzy Sets in Decision Analysis Operations Research and Statistics, Kluwer Academic Publishers, Massachusetts, 1998, pp. 349–387.
[28] R. Coppi, P. D'Urso, Regression analysis with fuzzy Informational Paradigm: a least-squares approach using membership function information, International Journal of Pure and Applied Mathematics 8 (2003) 279–306.
[29] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Classics in Applied Mathematics, vol. 15, SIAM, Philadelphia, PA, 1995.
[30] F.A. Graybill, An Introduction to Linear Statistical Models, vol. 1, McGraw-Hill, New York, 1961.
[31] H. Tanaka, S. Uejima, K. Asai, Linear regression analysis with fuzzy model, IEEE Transactions on Systems Man Cybernetics 12 (1982) 903–907.
[32] P. Guo, H. Tanaka, Dual models for possibilistic regression analysis, Computational Statistics & Data Analysis 51 (2006) 253–266.
[33] R. Coppi, M.A. Gil, H.A.L. Kiers, The fuzzy approach to statistical analysis, Computational Statistics & Data Analysis 51 (2006) 1–14.
[34] K.J. Kim, H. Moskovitz, M. Koksalan, Fuzzy versus statistical linear regression, European Journal of Operational Research 92 (1996) 417–434.
[35] M.L. Puri, D.A. Ralescu, Fuzzy random variables, Journal of Mathematical Analysis and Applications 114 (1986) 409–422.
[36] M. López-Díaz, D.A. Ralescu, Tools for fuzzy random variables: embeddings and measurabilities, Computational Statistics & Data Analysis 51 (2006) 109–114.

[37] H.T. Nguyen, B. Wu, Random and fuzzy sets in coarse data analysis, Computational Statistics & Data Analysis 51 (2006) 70–85.
[38] R. Viertl, Statistical Methods For Non-Precise Data, CRC Press, Boca Raton, FL, 1996.
[39] R. Coppi, M.A. Gil, H.A.L. Kiers (Eds.), Special Issue: The Fuzzy Approach to Statistical Analysis, Computational Statistics & Data Analysis 51 (2006) 1–451.
[40] M.L. Puri, D.A. Ralescu, The concept of normality for fuzzy random variables, Annals of Probability 13 (1985) 1373–1379.
[41] M. Montenegro, A. Colubi, M.R. Casals, M.A. Gil, Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable, Metrika 59 (2004) 31–49.
[42] G. González-Rodríguez, M. Montenegro, A. Colubi, M.A. Gil, Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data, Fuzzy Sets and Systems 157 (2006) 2608–2613.
[43] M.A. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, M.R. Casals, Bootstrap approach to the multi-sample test of means with imprecise data, Computational Statistics & Data Analysis 51 (2006) 148–162.
[44] P. D'Urso, P. Giordani, A weighted fuzzy c-means clustering model for symmetric fuzzy data, Computational Statistics and Data Analysis 50 (2006) 1496–1523.
[45] C. Dujet, N. Vincent, Force implication: a new approach to human reasoning, Fuzzy Sets and Systems 69 (1995) 53–63.
[46] N. Vincent, C. Dujet, A suggested conditional modus ponens, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5 (1997) 93–106.
[47] E. Benetto, C. Dujet, Uncertainty analysis of environmental impact assessment inferences: a possibilistic approach, in: Proceedings of the 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004), 2004, pp. 865-862.