Interval Discriminant Analysis: An Efficient Method to integrate Errors in Supervised Pattern Recognition

Philippe NIVLET
Institut Français du Pétrole
FRANCE
philippe.nivlet@ifp.fr

Frédérique FOURNIER
Institut Français du Pétrole
FRANCE
frederique.fournier@ifp.fr

Jean-Jacques ROYER
CNRS-CRPG
FRANCE
Jean-Jacques.Royer@ensg.inpl-nancy.fr

Abstract

In a statistical pattern recognition context, probabilistic algorithms like parametric or nonparametric discriminant analysis are designed to classify objects into predefined classes. Because these methods require precise input data, they cannot propagate uncertainties in the classifying process. In real case studies, this could lead to drastic misinterpretations of objects. We have thus developed an extension of these methods to directly propagate imprecise interval-form data. The computations are based on interval arithmetic, which appears to be an efficient tool to handle intervals. They consist in calculating successively interval conditional probability density functions and interval posterior probabilities, whose definitions are closely associated with the imprecise probability theory. The algorithms eventually assign any object to a subset of classes, consistent with the data and its imprecision. The resulting classifying model is thus less precise, but much more realistic than the standard one. The efficiency of this algorithm is tested on a synthetic case study.

Keywords. Discriminant analysis, interval arithmetic, imprecise probabilities.

1 Introduction

Supervised pattern recognition is a multivariate technique concerned with assigning objects to predefined categories, after observing some of their characteristics. It covers a large range of applications. For example, in Earth Sciences, and more particularly, in the area of reservoir characterisation for petroleum exploration and production, it is commonly applied to borehole data interpretation as also to seismic data analysis [2]. In a general context, supervised pattern recognition is first concerned with calibrating a function between the set of features and a set of classes with the help of a training sample for which both features and classes are known. This process is commonly referred to as “teaching the classifier”. This calibrated classifier is then used to to assign to new objects to the different classes.

In this context, discriminant analysis is a very powerful technique. Firstly, since it works in a probabilistic frame, probabilities of good assignment can be associated to the predicted categories. These probabilities are valuable for assessing the reliability of the interpretation. Secondly, discriminant analysis provides a guide for feature selection. This is very useful since in real situations, numerous attributes are generally measured. Criteria based on the performance of the discriminant function help in selecting the parameters that are the most relevant with respect to the prediction problem being addressed. Thirdly, discriminant analysis through nonparametric algorithms allows a proper identification of patterns, even if they are highly non linear, which is quite common in practice. Lastly, discriminant analysis can handle categorical features, or continuous ones, as it is the case in our paper.

Discriminant analysis, and more generally classical classifiers, consider that the analysed features are precise. However, in real life, this is never the case: data are usually imprecise, and may be in some case missing. This rises two natural questions:

- What are the consequences of this uncertainty on the final outcome of the assignment process?
- How can this uncertainty be propagated?

In this paper, we focus on the issue of propagating imprecision. All the considered features will be continuous and interval-form. To propagate these intervals, we need to define an extension of discriminant analysis, which we call interval discriminant analysis. This imprecise classifier is more general than most classifiers, as it relaxes the requirement of a single output

class. This relaxation is necessary, as the imprecision on the data gives rise to a set of probability distributions, instead just one. The definition of this interval classifier is thus intimately linked with the alteration from standard probability theory to imprecise probabilities [18], and more specifically, to credal sets [19,20].

Standard discriminant analysis will be presented in Section 2. The extension of this classifier to intervals, presented in Section 4, is based on interval arithmetic. (Section 3). Lastly, Section 5 shows a synthetic illustrative example.

2 Standard Discriminant Analysis

2.1 Algorithm

Let X be a random vector in IR^p and $C=\{C_1, ..., C_N\}$, a predefined set of classes. Discriminant analysis [6,7,8] aims at calibrating -and estimating the efficiency of- a statistical relationship between C and X:

$$C=R(X) \tag{1}$$

The classifying process is based on the Bayes rule, which estimates the posterior probability to assign an observation x to the class C_i:

$$p(C_i \mid x)=\frac{p(C_i).p(x \mid C_i)}{\sum_j p(C_j).p(x \mid C_j)} \tag{2}$$

with: $p(x \mid C_j)$, C_j's conditional probability density function (CPDF) ;
$p(C_j)$, C_j's prior probability.

In order to estimate the CPDFs $p(x \mid C_i)$ on the basis of the observation of the objects in each training subset C_i, parametric or nonparametric algorithms may be used. Formally, in the following, we will consider that each of the N training subsets C_i is composed of n_i objects x_{ij}. An object is characterized by p features, and will thus be written as $x_{ij}^t = (x_{ij}^{(1)} ; \dots ; x_{ij}^{(k)} ; \dots ; x_{ij}^{(p)})^t$.

The parametric algorithm consists in assuming that each training subset C_i is a random set from a Gaussian law, whose parameters (mean vector μ_i and variance-covariance matrix Σ_i) are estimated on the training sample (Equation 3).

$$p(x \mid C_i)=\frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)} \tag{3}$$

In that context, two approaches are possible.

Quadratic approach: the mean vectors and variance-covariance matrices are calculated independently on each training subset C_i. The boundaries between the classes obtained in IR^p are pieces of hyperboloids or hyperplanes, depending on the relative values of the variance-covariance matrices.
Linear approach: the variance-covariance matrices are supposed to be identical for each training subset related to C_j. The boundaries between the classes obtained in IR^p are pieces of hyperplanes.

This parametric approach may not work in real life problems, as variables are not usually Gaussian. In such a case, to hypothesize that these variables have a Gaussian distribution may appear somewhat unwarranted, and lead to erroneous conclusions. When the training sample size is sufficient, it is thus preferable to estimate the CPDFs with a nonparametric method. Silverman [17] reviews various approaches available in that scope. Among the most popular methods, is the k-nearest-neighbour method, in which the CPDFs are inversely proportional to the distance between x and its k^{th} nearest neighbour in IR^p. The method has two main drawbacks: the estimated CPDFs are highly non-differentiable and they have heavy tails. In this paper, we concentrate upon the kernel method for estimating the CPDFs:

$$p(x \mid C_i)=\frac{1}{n_i h^p}\sum_{j=1}^{n_i} K\left(\frac{x-x_{ij}}{h}\right) \tag{4}$$

with h smoothing parameter;
K kernel function.

A particularly attractive kernel shape, for minimizing the mean square error on the CPDF estimate, is given by Epanechnikov [5]:

$$\begin{cases} K(u)=\frac{p+2}{2N_p}(1-u^t u) \text{ if } |u|<1 \\ K(u)=0 \text{ otherwise} \end{cases} \tag{5}$$

with: N_p normalisation coefficient depending on p so that $\int K(u)du = 1$.

Equation 5 can be rewritten with the change in variable $u=\frac{x-x_{ij}}{h}$:

$$\begin{cases} K\left(\frac{x-x_{ij}}{h}\right)=\frac{p+2}{2N_p}\left(1-\left(\frac{x-x_{ij}}{h}\right)^t\left(\frac{x-x_{ij}}{h}\right)\right) \text{ if } \left|\frac{x-x_{ij}}{h}\right|<1 \\ K\left(\frac{x-x_{ij}}{h}\right)=0 \text{ otherwise} \end{cases} \tag{6}$$

or in an expanded form:

$$\begin{cases} K\left(\frac{x-x_{ij}}{h}\right)=\frac{p+2}{2N_p}\left(1-\sum_k\left(\frac{x^{(k)}-x_{ij}^{(k)}}{h}\right)^2\right) \\ \qquad\qquad \text{if } \left|\frac{x-x_{ij}}{h}\right|<1 \\ K\left(\frac{x-x_{ij}}{h}\right)=0 \text{ otherwise} \end{cases} \tag{7}$$

Once the different CPDFs and posterior probabilities are computed, the maximum likelihood rule is applied. It means that an observation x will be assigned to the class C_i which has maximum posterior probability $p(C_i|x)$.

2.2 Uncertainties

This standard discriminant analysis algorithm provides a first interpretation of the measurement data. Yet, it fails to account for data errors both in the calibration and in the assignment phases. The question we address in this paper is mainly linked with the propagation of the imprecision on the measurement x through the discriminant analysis process. Before presenting the way to do it, we summarise our motivations to model these by intervals.

When a measurement x is processed several times under the same experimental conditions with the same measuring tool, outcomes usually differ from one to another. These variations are usually explained by two main causes. First of all, the measuring tools are never perfectly precise. Moreover, even in such a situation, results would vary because of environmental condition changes between two measurements. These multiple outcomes are then used to infer the distribution of the quantity of interest. They may also be used to fit data with a given distribution. Within this latter context, errors on x are often assumed to have a Gaussian distribution. In a statistical pattern recognition for geophysical issues, this approach is not applicable [3], because measurements are almost never repeated. Thus, distributions cannot be inferred from multiple measurements. Intervals remain then the only way to model measurement imprecision. These intervals are usually provided by operators. They depend on the nominal precision of the measuring tool, and may sometimes be majored, when there is evidence of the bad quality of the experimental environment.

Once uncertainties are characterised, they need to be propagated. This question has been widely studied in the literature. Efron [4] has compared various algorithms based on the bootstrap principle to estimate the variance of the result. His point is to put emphasis on the errors due to the limited size of the calibration population. The Monte-Carlo approach is another way to propagate errors [16]. By generating many realisations of the initial data set and using them as inputs to the classical interpretation process, it is possible to estimate the output distribution. Yet, both Monte-Carlo and bootstrap methods are known to underestimate the errors on the final result [3].

In the present paper, we aim at bounding as tightly as possible the errors on the final results of discriminant analysis due to errors on the measurements x_i of the training sample. This means to compute Equations 2 to 6, by replacing real observations x_{ij} by interval observations $[x_{ij}^-; x_{ij}^+]$. Interval arithmetic, whose principles are explained below, allows interval-based uncertainties to be handled.

3 Interval Arithmetic: Basic Concepts

Interval arithmetic was first developed in [14], to compute interval data. Recent developments of this theory can be found in [10]. By convention, in the following, I(IR) will designate the set of real intervals, and $I(IR)^p$ the set of real p-dimension arrays of intervals (also called p-dimensional pavements). The minimum of any real interval $x_{[]}$ will be quoted x^-, and its maximum, x^+.

In this section, the main interval computation properties are defined, beginning with the four standard arithmetic operations.

Definition 1. Let $\boldsymbol{x}=[x^-; x^+]$ and $\boldsymbol{y}=[y^-; y^+]$ be two intervals in I(IR). Then,

$$\begin{cases} x_{[]}+y_{[]}=[x^-+y^-; x^++y^+] \\ x_{[]}-y_{[]}=[x^--y^+; x^+-y^-] \\ x_{[]}.y_{[]}=[\min\{x^-y^-; x^-y^+; x^+y^-; x^+y^+\}; \\ \qquad \max\{x^-y^-; x^-y^+; x^+y^-; x^+y^+\}] \\ \frac{1}{x_{[]}}=\left[\frac{1}{x^+}; \frac{1}{x^-}\right] \text{if } 0\notin x_{[]} \\ \frac{y_{[]}}{x_{[]}}=y_{[]}.\frac{1}{x_{[]}} \text{if } 0\notin x_{[]} \end{cases} \tag{8}$$

We also need the following to define the extension of comparison operators to intervals.

Definition 2. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two intervals in I(IR), then

$$x_{[]} \succ y_{[]} \Leftrightarrow x^- > y^+ \tag{9}$$

These definitions are the basis for more complex computations defined below.

Definition 3. The inclusion function $f_{[]}$ of any real function f of p real variables is defined as follows:

$$\begin{aligned} f_{[\,]} &: \mathrm{I(IR)}^p \rightarrow \mathrm{I(IR)} \\ & x_{[\,]} \mapsto y_{[\,]} \supseteq \{ y = f(x) \in \mathrm{IR} \mid x \in \boldsymbol{x} \} \end{aligned} \tag{10}$$

It is said to be optimal when $f_{[\,]}(x_{[\,]}) = \{ y = f(x) \in \mathrm{IR} \mid x \in x_{[\,]} \}$.

For elementary functions, as the exponential, the optimal inclusion function is easy to find: $exp_{[\,]}(\boldsymbol{x}) = [\exp(x^-); \exp(x^+)]$. It is not the case for more complex ones. Yet, combining elementary interval function definitions given by Equation 10 with the basic arithmetic operations (Equation 8), it is still possible to compute the interval extension of any function. This interval extension is called the natural inclusion function. As pointed out by Alefeld and Herzberger [1], the width of this interval function is generally overestimated, especially when variables appear several times in the equation, as it is the case in the following simple example: Let $f_{[\,]}(x_{[\,]}) = x_{[\,]}.x_{[\,]}$; then, after Equation 8, $f_{[\,]}([-1;1]) = [-1;1]$ which is true in the sense of Equation 10, but far from optimal. This is caused by the fact that interval arithmetic computes each occurrence of a single interval variable $x_{[\,]}$, as if it were an independent variable. To overcome this drawback, the analytical expression of the function f has to be transformed, if possible, to avoid the redundant variables in its mathematical formulation. For example, the previous quadratic function would return the optimal bounds if it were written $f_{[\,]}(x_{[\,]}) = x_{[\,]}^2$.

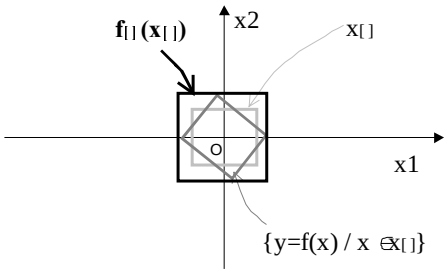

Figure 1: Illustration of the wrapping effect

Another source of overbounding in interval computations is known as the wrapping effect. This effect is illustrated in Figure 1 in the case where f is a rotation of center O. As interval computations can only generate pavements, which edges are parallel to the main axes x1 and x2, they necessarily generate overestimations.

In order to reduce both of these undesirable overbounding effects, it is possible to use a "branch-and-bound"-like algorithm [9], whose principles are illustrated on Figure 2. This kind of iterative algorithm was first developed in the context of global optimization [11].

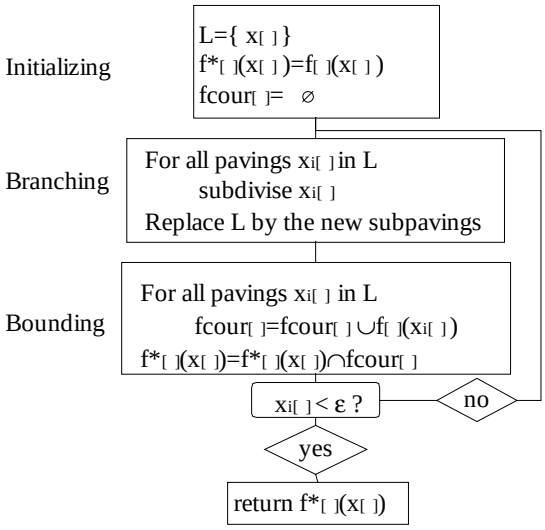

Figure 2: "branch-and-bound" algorithm

As shown in Figure 2, the "branch-and-bound" algorithm allows to compute successive bounds to any interval function $f^*_{[\,]}$ which have decreasing width, by subpaving more and more closely the initial pavement $x_{[\,]}$. Although this algorithm is quite popular for sake of simplicity, it does not converge very quickly.

These concepts will be applied to Equations 2 to 6 to integrate errors in variables in the discriminant analysis model. The integration of uncertainties leads to compute interval probabilities.

4 Interval Discriminant Analysis

In order to build an interval discriminant analysis algorithm, we have to extend the three main steps in the standard algorithm to intervals:

- Computation of interval CPDFs;
- Computation of interval posterior probabilities;
- Building the interval classifier.

It could be argued that one could have used standard optimization methods [13] to compute bounds from these intervals. Yet, in many real life problems, this solution is inconceivable, because of its computational cost. Interval arithmetic thus appears as a more feasible alternative to propagate uncertainty.

4.1 Interval Probability Density Functions

4.1.1 Nonparametric approach

To build an interval arithmetic based nonparametric discriminant analysis, we first have to compute an

interval extension for Equation 6, which is a weighted sum of quadratic terms. Each term is easily extended to intervals, using basic interval arithmetic definitions (Equation 8). Figure 3 shows the resulting optimal interval kernel function when only one attribute is taken under consideration (p=1).

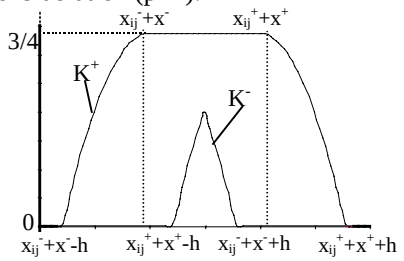

Figure 3: The interval Epanechnikov kernel (p=1)

In Figure 3, it must be noticed that the upper and lower bounds depend directly on h, and on the uncertainties on observations x and x_{ij}. In some cases (when $x_{ij}^{+}+x^{+}-h>x_{ij}^{-}+x^{-}+h$), the lower bound can be equal to the null function. Imprecision on the kernel interval estimate, and thus on the interval CPDF, is then maximum.

When p>1, using basic interval arithmetic definitions would lead to overestimate the interval kernel, because of the constraint $\left|\frac{x-x_{ij}}{h}\right|<1$ in Equation 6, which expresses a dependency between each component $x_{ij}^{(k)}$. To overcome this problem, we have to recall that both bounds of the multivariate interval kernel must remain positive. Integrating this simple condition is sufficient to get the optimal multivariate kernel inclusion function $K_{[\,]}\left(\frac{\boldsymbol{x}_{[\,]}-\boldsymbol{x}_{ij[\,]}}{h}\right)$.

The CPDF calculations are then straightforward, since Equation 3 is a sum of independent kernel functions. The interval extensions of the CPDFs will then be:

$$p_{[\,]}\left(\boldsymbol{x}_{[\,]}\middle|\,C_i\right)=\frac{1}{n_i h^p}\sum_j K_{[\,]}\left(\frac{\boldsymbol{x}_{[\,]}-\boldsymbol{x}_{ij[\,]}}{h}\right) \qquad (11)$$

Equation 11 outcomes the optimal interval extension of the nonparametric CPDFs, in the sense that the resulting interval is as tight as possible.

4.1.2 Parametric Approach

To extend the Gaussian CPDF (Equation 3) to intervals implies to compute the inclusion function:

$$p_{[\,]}\left(\boldsymbol{x}_{[\,]}\middle|\,C_i\right)=\frac{1}{(2\pi)^{p/2}\left|\Sigma_{i[\,]}\right|^{1/2}}e^{-\frac{1}{2}\left(\boldsymbol{x}_{[\,]}-\boldsymbol{\mu}_{i[\,]}\right)^t\Sigma_{i[\,]}^{-1}\left(\boldsymbol{x}_{[\,]}-\boldsymbol{\mu}_{i[\,]}\right)} \qquad (12)$$

When p=1, the extension of a Gaussian CPDF is straightforward to compute, as in the nonparametric case, once the mean $\mu_{i[\,]}$ and the variance $\sigma^2_{i[\,]}$ have been enclosed (Figure 4). Bounds of $\mu_{i[\,]}$ are computed with the basic interval arithmetic rules (Equation 8). In order to estimate the bounds of $\sigma^2_{i[\,]}$, simple optimization procedures (as the conjugate gradient method) are used to solve the following constrained problems:

$$\begin{cases}\sigma_i^{2-}=\underset{x_{ij}\in x_{ij[\,]},\forall j}{\text{Min}}\left\{\frac{1}{n_i-1}\sum_j\left(x_{ij}-\frac{1}{n_i}\sum_k x_{ik}\right)^2\right\}\\ \sigma_i^{2+}=\underset{x_{ij}\in x_{ij[\,]},\forall j}{\text{Max}}\left\{\frac{1}{n_i-1}\sum_j\left(x_{ij}-\frac{1}{n_i}\sum_k x_{ik}\right)^2\right\}\end{cases} \qquad (13)$$

These methods are local optimization algorithms. However, in the particular case of Equations 13, because the variance is a convex function with respect to the variables x_{ij}, the local minimum (or maximum), is also a global minimum (or maximum). These methods thus outcome the exact interval variance, given the imprecision on the data.

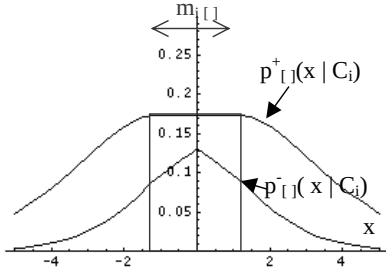

Figure 4: The interval Gaussian distribution (p=1)

Note that this computation of this interval Gaussian CPDF is nevertheless not optimal. The overestimation of the interval CPDFs results from the fact that we have considered independently the means and variances in the interval computations. Formally, this is not true and the computed CPDFs are only approximations.

In the multivariate case (p>1), the extension of a Gaussian distribution is more complex. The natural inclusion function of Equation 12 usually leads to compute very wide intervals. This overbounding is due to the wrapping effect. It is amplified as the expression $(x-\mu_i)$ occurs several times in Equation 3. To minimize these effects, the following algorithm was developed:

1- Computation of the p(p+1)/2 distinct interval terms of the variance-covariance matrix Σ_i using a conjugate gradient method to find the minimum and maximum of the experimental covariances :

$$\frac{1}{n_i - 1}\sum_j \left(x_{ij}^{(l)} - \frac{1}{n_i}\sum_k x_{ik}^{(l)} \right)\left(x_{ij}^{(m)} - \frac{1}{n_i}\sum_k x_{ik}^{(m)} \right),$$ for all indexes l and m. This first step allows to compute the interval experimental variance-covariance matrices $\Sigma_{i[\,]}$, which is optimal for the same reason as in the univariate case.

2- "Diagonalization" of the interval $\Sigma_{i[\,]}$ matrix. This diagonalization is achieved in two substeps: The interval matrix is first pre- and post- multiplied by the rotation matrix R_θ which would diagonalize the real matrix $(\Sigma_i^- + \Sigma_i^+)/2$. This preconditioning step, known as the interval Jacobi method [15] transforms $\Sigma_{i[\,]}$ into an interval matrix $\Sigma'_{i[\,]}$ whose non-diagonal terms have minimal width. The interval matrix $\Sigma'_{i[\,]}$ is then replaced by an including interval diagonal matrix $\Sigma''_{i[\,]}$ whose out-diagonal terms are null. This operation necessarily amplifies the width of diagonal terms. Finally, obtaining $\Sigma''_{i[\,]}$ is equivalent to searching the set of eigenvalues of $\Sigma_{i[\,]}$, but considering that the eigenvectors of this imprecise matrix are known.

At the end of this step, problem formulated by Equation 12 is replaced by finding the interval function $p'_{[\,]}$.

$$p'_{[\,]}(\boldsymbol{x}_{[\,]} \mid C_i) \supseteq p_{[\,]}(\boldsymbol{x}_{[\,]} \mid C_i) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}_{[\,]} - \boldsymbol{\mu}_{i[\,]})^t R_\theta{}^t \Sigma''^{-1}_{i[\,]} R_\theta (\boldsymbol{x}_{[\,]} - \boldsymbol{\mu}_{i[\,]})}}{(2\pi)^{p/2} \left| \Sigma''_{i[\,]} \right|^{1/2}} \quad (14)$$

Step 2 is the major source of approximation in the interval computations, mainly because we have explained the imprecision on the interval variance-covariance matrix by the imprecision on eigenvalues computed in a fixed basis, which is a simplification of the problem.

3- Because of the rotation matrix R_θ, the direct application of interval natural inclusion to Equation 14 is not optimal. We thus have to refine the inclusion function by subpaving $(\boldsymbol{x}_{I\,J} - \boldsymbol{\mu}_{i\,J})$ into n_s disjoint subdomains $\boldsymbol{X}_{k\,J}$. By using the same "branch-and-bound" algorithm principle, we compute on each of these n_s pavements an interval natural inclusion function $p_{k[\,]}(\boldsymbol{x}_{I\,J} \mid C_i,\ \boldsymbol{x} - \boldsymbol{\mu}_i \in \boldsymbol{X}_{k\,J})$. The CPDF interval is then computed as:

$$p'_{[\,]}(\boldsymbol{x}_{[\,]} \mid C_i) = \bigcup_k p'_{k[\,]}(\boldsymbol{x}_{[\,]} \mid C_i, \boldsymbol{x}_{[\,]} - \mu_{i[\,]} \in \boldsymbol{X}_{k[\,]}) \quad (15)$$

These interval functions are not optimal inclusions of Gaussian CPDFs. However, they are tighter enclosures of these functions, and are thus preferable to the interval natural inclusion functions.

4.2 Interval posterior probabilities

Once the interval CPDFs have been computed, it is straightforward to extend the Bayes rule to interval probabilities. To obtain directly the optimal enclosure, Equation 2 must be rewritten as:

$$p(C_i \mid x) = \left(1 + \sum_{j \neq i} \frac{p(C_j)}{p(C_i)} \frac{p(x \mid C_j)}{p(x \mid C_i)} \right)^{-1} \quad (16)$$

Using the interval arithmetic basic rules, the interval posterior probability optimal bounding functions are then:

$$\begin{cases} p^-(C_i \mid \boldsymbol{x}_{[\,]}) = \left(1 + \sum_{j \neq i} \frac{p(C_j)^+ p(\boldsymbol{x}_{[\,]} \mid C_j)^+}{p(C_i)^- p(\boldsymbol{x}_{[\,]} \mid C_i)^-} \right)^{-1} \\ p^+(C_i \mid \boldsymbol{x}_{[\,]}) = \left(1 + \sum_{j \neq i} \frac{p(C_j)^- p(\boldsymbol{x}_{[\,]} \mid C_j)^-}{p(C_i)^+ p(\boldsymbol{x}_{[\,]} \mid C_i)^+} \right)^{-1} \end{cases} \quad (17)$$

These interval posterior probabilities $p_{[\,]}(C_i \mid \boldsymbol{x}_{I\,J})$ may be seen as a particular application of credal sets [12,19,20], which are defined as convex sets of probability distributions bounded by linear constraints.

4.3 Interval classifier

The final step of the classifying procedure consists in assigning the imprecise object $\boldsymbol{x}_{I\,J}$ to a subset of possible classes. Two criteria based on the interval extension of the maximum likelihood rule are defined.

The first one is based on the direct comparison of the different $p_{[\,]}(C_i \mid \boldsymbol{x}_{I\,J})$, i=1…N. We have to find which of these intervals are dominated, by using the interval extension of comparison (Definition 2). The subset of dominated classes is then discarded from the set of possible output classes. The interval posterior probabilities of the remaining undominated classes thus necessarily overlap. This subset is the final outcome of the classifying procedure. This criterion is also known in the literature as strong dominance criterion [20].

However, as we have stated in Section 2, $p_{[\,]}(C_i \mid \boldsymbol{x}_{I\,J})$ may also be seen as bounds to sets of posterior probabilities, when the CPDFs vary in the intervals $p_{[\,]}(\boldsymbol{x}_{I\,J} \mid C_i)$, and the prior probabilities, in the intervals $p_{[\,]}(C_i)$. In this case, the strong dominance criterion necessarily overestimates

the assignment uncertainty. This overestimation may be explained by considering that, as all the interval probabilities depend on the same quantities $p_{[\,]}(\boldsymbol{x}_{//} \mid C_i).p_{[\,]}(C_i)$, they are not independent.

A less pessimistic criterion to discard the dominated classes is based on the direct comparison of the intervals $p_{[\,]}(\boldsymbol{x}_{//} \mid C_i)p_{[\,]}(C_i)$, i=1…N. Before doing so, we sort the values of $p^+(\boldsymbol{x}_{//} \mid C_i)p^+(C_i)$ by decreasing order:

$$p^+\left(\boldsymbol{x}_{[\,]} \middle| C_{i_1}\right)p^+\left(C_{i_1}\right) \geq \ldots \geq p^+\left(\boldsymbol{x}_{[\,]} \middle| C_{i_N}\right)p^+\left(C_{i_N}\right) \qquad (18)$$

Then, using the basic definition for comparisons of intervals (Equation 10), it holds:
If $p^-\left(\boldsymbol{x}_{[\,]} \middle| C_{i_1}\right)p^-\left(C_{i_1}\right) \geq p^+\left(\boldsymbol{x}_{[\,]} \middle| C_{i_N}\right)p^+\left(C_{i_N}\right)$, $\boldsymbol{x}_{//}$ is assigned into C_{i_1}. Otherwise, the algorithm cannot assign among C_{i_1} and C_{i_2}, and we repeat the former test with C_{i_1} and C_{i_3}; … ; C_{i_1} and $C_{i_{\{p\}}}$, until $p^+\left(\boldsymbol{x}_{[\,]} \middle| C_{i_1}\right)p^+\left(C_{i_1}\right) \geq p^+\left(\boldsymbol{x}_{[\,]} \middle| C_{i_{\{p\}}}\right)p^+\left(C_{i_{\{p\}}}\right)$.

This criterion is the credal dominance criterion as defined in [20], in the case where CPDFs are bounded by pavements. It is optimal when the quantities $p_{[\,]}(\boldsymbol{x}_{//} \mid C_i).p_{[\,]}(C_i)$ are independent. This is actually the case when interval CPDFs are estimated by the nonparametric algorithm, or the quadratic approach of the parametric algorithm. In the linear approach, it is not. The credal criterion then overestimates the assignment uncertainty. To improve it, the last remaining solution is to use a branch-and-bound algorithm, as defined in Section 2.

As a conclusion, the imprecisions on the observations of the training sample have been propagated in a discriminant analysis algorithm. These imprecisions on the data points generate uncertainties on the CPDFs and on the posterior probabilities, which cause uncertain assignments.

5 Case Study

In order to illustrate these theoretic developments of interval discriminant analysis, we show a very simple synthetic case study. In this example, we want to calibrate an interval classifying function between objects defined by two features x and y and three predefined classes. To train the classifier, 100 data per class are available. Figure 5 shows that the three training classes are well separated. This usually implies that the two features are discriminant enough to build a good-quality classifier.

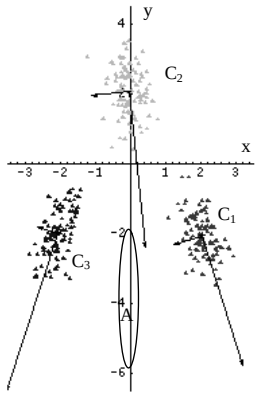

Figure 5: Training sample in the plane (x; y)

In this section, we will only show the results obtained by the quadratic approach to standard and interval discriminant analyses. For the interval algorithm, we will consider that the imprecision is ±0.07 on the data's first feature, and ±0.11 on the data's second feature.

Figure 6 shows the calibrated function C(x; y) obtained by standard discriminant analysis. The plane (x; y) is then simply splitted into three parts, separated by hyperbolic boundaries. It is consequently possible to assign any object (x, y) to a specific class. Zones characteristic of each training class are also circled and dashed on Figure 6.

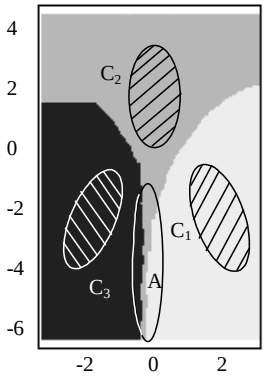

Figure 6: Calibrated classifying function in the plane (x;y)

However, because of the requirement of precision of this standard approach, the assignment may appear somewhat arbitrary, especially in zones of the plane (x; y) near the boundaries between each class. Moreover, in the region named A, the calibrated function predicts class C_2, although on Figure 1, it appears to be much closer from the training subsets C_1 and C_3.

By relaxing this precision requirement, an interval classifier is calibrated. Table 1 summarises the interval means and variance-covariance matrices of the different classes.

C_i	$\mu_{i[\,]}$	$\Sigma_{i[\,]}$
1	$\begin{pmatrix} [1.99;2.13] \\ [-2.17;-1.95] \end{pmatrix}$	$\begin{pmatrix} [0.093;0.130] & [-0.118;-0.054] \\ [-0.118;-0.054] & [0.310;0.418] \end{pmatrix}$
2	$\begin{pmatrix} [-0.10;0.04] \\ [1.95;2.17] \end{pmatrix}$	$\begin{pmatrix} [0.093;0.130] & [-0.067;0.000] \\ [-0.067;0.000] & [0.387;0.507] \end{pmatrix}$
3	$\begin{pmatrix} [-2.11;-1.97] \\ [-2.10;-1.88] \end{pmatrix}$	$\begin{pmatrix} [0.086;0.123] & [0.090;0.158] \\ [0.090;0.158] & [0.378;0.497] \end{pmatrix}$

Table 1: Summary of interval means and variance-covariance matrices for each class

The procedure described in Section 4.1.2 is then applied in order to estimate the associated interval diagonal $\Sigma''_{i[\,]}$ and rotation R_θ matrices (Table 2).

C_i	R_θ	$\Sigma''_{i[\,]}$
1	$\begin{pmatrix} 0{,}955 & -0{,}294 \\ 0{,}294 & 0{,}955 \end{pmatrix}$	$\begin{pmatrix} [0.031;0{,}146] & 0 \\ 0 & [0.216;0.536] \end{pmatrix}$
2	$\begin{pmatrix} 0{,}995 & -0{,}099 \\ 0{,}099 & 0{,}995 \end{pmatrix}$	$\begin{pmatrix} [0.082;0.139] & 0 \\ 0 & [0.363;0.517] \end{pmatrix}$
3	$\begin{pmatrix} 0{,}949 & 0{,}314 \\ -0{,}314 & 0{,}949 \end{pmatrix}$	$\begin{pmatrix} [0.012;0.123] & 0 \\ 0 & [0.231;0.645] \end{pmatrix}$

Table 2: Rotation matrices and interval eigenvalues for each class

In order to improve the estimation of the interval CPDFs, each initial mean pavement was parted into 20 subdomains.

To check the efficiency of this subpaving approach, we show the step-by-step computations involved in the assignment of an object $\boldsymbol{x}^*_{//}$=([0.93,1.07];[-0.11;0.11]). Table 3 illustrates the sharpening of the interval CPDFs obtained by using these procedures. For each class, it compares the quantities $p_{[\,]}(\boldsymbol{x}^*_{//} \mid C_i)$, obtained with the direct interval natural extension IE (Equation 12), with Equation 14 (IE2), and after refining the IE2 approximation by subpaving the initial interval mean pavements into 20 subdomains (IE3). In Table 3, the results provided by the first two algorithms (IE and IE2) are not relevant to our problem, because the width of the output intervals is very large, and the three interval CPDFs intervals overlap. However, IE2 provides tighter enclosures of the CPDFs. At least, IE3 improves tremendously the accuracy of the estimation.

C_i	IE	IE2	IE3
1	$[0;+\infty[$	$[2.10^{-38};1.10^{6}]$	$[1.10^{-7};6.10^{-3}]$
2	$[8.10^{-37};3.10^{-1}]$	$[9.10^{-11};8.10^{-2}]$	$[3.10^{-5};3.10^{-3}]$
3	$[0;+\infty[$	$[1.10^{-461};3.10^{74}]$	$[1.10^{-105};3.10^{-13}]$

Table 3: Comparison of three interval estimations of $p_{[\,]}\{\boldsymbol{x}^*_{//} \mid C_i\}$

Table 4 reports the resulting interval posterior probabilities, computed with the generalised Bayes rule. As it could be expected, both IE and IE2 generate non-informative posterior probabilities [0;1] for all the classes. IE3 outcomes more precise results, as $p(C_3 \mid \boldsymbol{x}^*_{//})$ is null. As a result, IE and IE3 cannot assign the object among the three classes, whereas IE3 only produces the outcome $\{C_1, C_2\}$. Note that for this object, the results are the same with both dominance criteria.

C_i	IE	IE2	IE3
1	[0;1]	[0;1]	[0.01;0.99]
2	[0;1]	[0;1]	[0.01;0.99]
3	[0;1]	[0;1]	[0;0]

Table 4: Comparison of three interval estimations of $p_{[\,]}\{C_i \mid \boldsymbol{x}^*_{//}$

After illustrating the propagation of imprecision on the assignment of a particular object, it is worth examining the general results of algorithm IE3, which appear much more efficient than the two others. Figure 7 maps the possible assigned classes in the attribute space (x;y) . For example, the white zone correspond to objects which could not be assigned among the three classes.

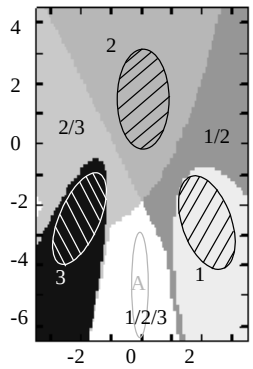

Figure 7: Calibrated imprecise classifying function in the plane (x; y)

We notice that regions where the assignment is uncertain correspond to the boundary regions between the 3 classes shown on Figure 6. This statement is a clear evidence that in these zones, the assignment between two or more classes is not evident. Furthermore, in region A, the interval classifier is not able to assign between any class. Usually, this situation appears for objects which are very different from the training population. In this case, very low CPDFs are computed, associated with a non-negligible level of uncertainty.

As a conclusion, the imprecise classifying model built in this synthetic case study appears to be much more realistic than the standard one. The use of intervals also prevents us from predicting arbitrary precise class for objects which cannot be discriminated between the predefined classes.

6 Conclusion

In many real-life issues, classifying is a strong need to make the measured features more valuable. So far, discriminant analysis was proved to be an efficient tool to achieve this purpose. Yet, this statistical method fails to assess the imprecision on the interpreted model due to the measurement errors.

This paper has shown a direct application of interval analysis to solve this problem. More specifically, it gives a quantitative assessment of the stability of predicted class, given by the standard algorithm. It also requires less prior knowledge on the studied objects: for example, it does not require any knowledge on the exact distribution of measurement errors, as the Monte-Carlo methods do. Moreover, the solution given by interval analysis is always reliable, because of the inclusion property. As a consequence, it provides a much more realistic interpretative model of the studied objects.

References

[1] G. Alefeld, and J. Herzberger. *Introduction to Interval Computations*, Academic Press, New York. 1983.

[2] P. Y. Dequirez, F. Fournier, T. Feuchtwanger, D. Torriero. Integrated Stratigraphic and Lithologic Interpretation of the East-Senlac Heavy-Oil Pool. *Expanded Abstracts of the 65th Annual International SEG Meeting*. 104 - 107. 1995.

[3] D. I. Doser, K. D. Crain, M. R. Baker, V. Kreinovich, M. C. Gerstenberger. Estimating Uncertainties for Geophysical Tomography. *Reliable Computing*, 4:241 - 268. 1998.

[4] B. Efron. Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. *Biometrika*. 68 – 3:589 - 599. 1981.

[5] V. A. Epanechnikov. Nonparametric Estimate of a Multivariate Probability Density: *Theor. Probab. Appl.*. 14:179 - 188. 1969.

[6] R. A. Fisher. The Use of Multiple Measurements in Taxinomic Problems. *Annals of Eugenics*. 7:153 - 158. 1936.

[7] K. Fukunaga. *Introduction to Statistical Pattern Analysis*. Academic Press, New York. 1972

[8] D. J. Hand. *Discrimination and Classification.* Eds J. Wiley and Sons, New York. 1981.

[9] E. R. Hansen. *Global Optimization using Interval Analysis*. Marcel Dekker, New York. 1992

[10] L. Jaulin, J. L. Boimond, and H. Hardouin. Estimation via Set Inversion: Application to Discrete-Event Systems. Reliable Computing 165 – 173. 1999.

[11] R. Baker Kearfott. In *Interval Arithmetic Methods for Non Linear Systems and Non Linear Optimization; Introduction and Status* Publications in Operations Research, Holland. 533-542. 1992.

[12] I. Levi. *The Enterprise of Knowledge.* MIT Press, London. 1980.

[13] D. G. Luenberger. *Optimization by Vector Space.* John Wiley & Sons, New York. 1969.

[14] R. E. Moore. *Interval Analysis*. Prenctice-Hall, Englewood Cliffs. 1969.

[15] R. Pavec. Some Algorithms Providing Rigorous Bounds for the Eigenvalues of a Matrix. *Journal of Universal Computer Science*. 1 – 7:548 – 559. 1995.

[16] B. D. Ripley., *Stochastic Simulation.* John Wiley and Sons, New York. 1987.

[17] B. W. Silverman. *Density Estimation for Statistical and Data Analysis*. Chapman and Hall, London. 1986.

[18] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London. 1991.

[19] M. Zaffalon. A Credal Approach to Naive Classification: *Proceedings of the 1st International Sympsosium of Imprecise Probabilities and Their Applications*. 1999.

[20] M. Zaffalon. The Naive Credal Classifier: *Journal of Statistical Planning and Inference*. 2000 (in press).