

# A criterion for the comparison of binary classifiers based on a stochastic dominance with an application to the sale of home insurances

María Concepción López-Díaz<sup>a</sup>, Miguel López-Díaz<sup>b,\*</sup>, Sergio Martínez-Fernández<sup>c</sup>

<sup>a</sup>Departamento de Matemáticas, Universidad de Oviedo. C/Federico García Lorca 18. E-33007 Oviedo, Spain. e-mail: cld@uniovi.es

<sup>b</sup>Departamento de Estadística e I.O. y D.M., Universidad de Oviedo. C/Federico García Lorca 18. E-33007 Oviedo, Spain. e-mail: mld@uniovi.es

<sup>c</sup>Sección de Parámetros de Riesgo, Liberbank. Plaza de La Escandalera 2. E-33003 Oviedo, Spain. e-mail: smartinezf@liberbank.es and Departamento de Estadística e I.O. y D.M., Universidad de Oviedo. C/Federico García Lorca 18. E-33007 Oviedo, Spain. e-mail: sergiofernandez@uniovi.es

## Abstract

Binary classification is an essential matter in multiple real-life problems, and so, the comparison of the performance of classifiers is a key issue. A criterion for that purpose is introduced in this manuscript. That criterion is based on a stochastic dominance, and permits to compare classifiers in subgroups of the population with the same size. By means of the new criterion, the alteration of the size of the subgroups where classifiers are compared, does not entail the modification of the suitable classifier. Characterization results of the criterion are proved. For that purpose, connections of the criterion with the theory of copulas, and with a tool introduced in the

---

\*Corresponding author

manuscript, the so-called continuity modelling vector, are essential. An application to the comparison of some classifiers for the detection of purchasers of home insurances is developed.

Keywords: alternative accumulated improvement curve, binary classifier, continuity modelling vector, copula, insurance, target

## 1 Introduction

Classification is an appealing topic because of its importance in multiple fields, like insurance, banking, biological classification, medical diagnosis, computer vision, internet search engines, etc. Very basically, classification aims to designate to which of a group of categories an individual belongs. A classification procedure is said to be binary if the number of categories is equal to two. Research on this kind of classification has experienced a significant growth in recent years. For instance, binary classification is a key issue for commercial purposes in marketing strategies, especially in the development of marketing programs to launch campaigns among clients, or in the right classification of patients with respect to an illness, or in the identification of possible churn attrition, or possible credit card fraud, etc.

Binary classification systems are frequently carried out by means of a bidimensional random vector (modelling vector), whose components are the so-called classifier and target. A target is a Bernoulli random variable with parameter  $q$ , where  $q$  denotes the true and unknown proportion of individuals which belong to an specific category. For instance, when an insurance company analyzes which potential clients will purchase a life insurance,  $q$  is the true proportion of them which will acquire such an insurance. The target variable takes on value 1 at an individual which will purchase the insurance, otherwise zero. The value the target assigns to each individual is unknown, and the aim of any binary classification system is the prediction of that. Such an estimation is performed with the information provided by the value of the classifier at each individual.

Multiple techniques have been proposed and designed to develop classifiers, using, for example, neuronal networks, decision trees, random forests, logistic regression, Bayesian analysis, mesh methods, etc. (see, for instance, Breiman (2001), Wei and Chiu (2002), Hwang *et al.* (2004), Buckinx and Van den Poel (2005), Hung *et al.* (2006), Qi *et al.* (2009), Figini and Giudici (2010), Günther *et al.* (2014), etc.).

A key matter is the comparison of classifiers. Identifying suitable classifiers to estimate a target is a fundamental issue because of the importance of underlying problems. Special attention has been focused on this question during the last years (see, for instance, Lloyd (1998), Lee (1999), Hand (2009), Hand and Zhou (2009), Hand (2010), Hand (2012), Hand and Anagnostopoulos (2012), Hand and Anagnostopoulos (2013), Yousef (2013), etc.)

A criterion for the comparison of classifiers based on the theory of stochastic orders was introduced in López-Díaz *et al.* (2017). We describe that procedure briefly, in order to clarify the proposal of the present manuscript.

Consider a modelling vector  $X = (C_X, T_X)$ , where the target  $T_X$  follows a Bernoulli distribution with parameter  $q$ , and  $C_X$  is the classifier. The mapping  $M_X : (0, 1) \rightarrow \mathbb{R}$ , with  $M_X(p) = P(T_X = 1/C_X \geq F_{C_X}^{-1}(1 - p))$  for any  $p \in (0, 1)$ , is said to be the accumulated improvement curve of the modelling vector  $X$ . Given any  $p \in (0, 1)$ , we consider the smallest group containing at least the  $100p\%$  of the individuals with the largest values of the classifier  $C_X$ , and in such a subgroup, we study the probability that the target takes on value 1. That represents the probability of being right when we estimate the value of the target as 1 at those individuals.

Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors where  $T_X$  and  $T_Y$  follow Bernoulli distribution with parameter  $q$  (from an applied point of view, usually  $T_X = T_Y$ ). It will be said that the modelling vector  $X$  is less than  $Y$  in the accumulated improvement curve stochastic order, if  $M_X(p) \leq M_Y(p)$  for any  $p \in (0, 1)$ . It will be denoted by  $X \preceq_M Y$ . Roughly speaking, the relation  $X \preceq_M Y$  means that for any  $p \in (0, 1)$ , given the  $100p\%$  of the individuals with the largest values of the corresponding classifiers  $C_X$  and  $C_Y$ , the probability of carrying out right classifications by estimating as 1 the value of the target at those individuals, is greater (at least the same) with classifier  $C_Y$  than with classifier  $C_X$ .

It is interesting to remark that rating systems of classifiers based on a single numeric value, by means of a global assessment of classifiers, like ROC and CAP indexes (areas under ROC and CAP curves, respectively), lead to a total preorder, and so permit to order any two classifiers. However, a rating could fail when those classifiers are applied in subgroups of the whole population, like those given by  $C_X \geq F_{C_X}^{-1}(1 - p)$  and  $C_Y \geq F_{C_Y}^{-1}(1 - p)$ , for a specific value  $p \in (0, 1)$ . The procedure based on the accumulated improvement curve avoids that important drawback, considering the comparison in all those groups. A key advantage of that technique is that when  $X \preceq_M Y$ , the modification of the sizes of the groups, that is, the change of

$p$ , does not entail the modification of the suitable classifier. The reader is referred to López-Díaz *et al.* (2017) for a detailed theoretical analysis of the accumulated improvement curve criterion, and an application of that to the customer attrition problem in commercial banking.

A critique which could be assigned to the above procedure is that when discrete classifiers are considered, the subgroups of the population given by  $C_X \geq F_{C_X}^{-1}(1 - p)$  and  $C_Y \geq F_{C_Y}^{-1}(1 - p)$  with  $p \in (0, 1)$ , could have quite different sizes. The comparison of classifiers in subgroups which are not similar in size is commonly avoided. We could say that the root cause of this piece of work is based on how to solve that drawback of the above criterion for the rating of classifiers. The criterion introduced in the present manuscript avoids the above inconvenient, considering groups with the same size in the comparison of classifiers, for all possible sizes. The main contributions of the paper in the different sections are the following. The new mathematical model to compare classifiers is given in Section 3. Its definition is based on the comparison of the so-called alternative accumulated improvement curves. Some preliminaries results are developed in this section. Section 4 is essential to analyze the proposed model from a theoretical point of view. It is proved the existence of, what we have coined, the continuity modelling vector, that is, given any modelling vector, there exists another modelling vector with the same alternative accumulated improvement curve, and so at the same level in the new criterion, whose classifier has uniform distribution on the interval  $(0, 1)$ . This result is key to obtain posterior mathematical properties, since the use of continuous classifiers simplifies mathematical developments considerably. Section 5 is devoted to prove characterization results of the new criterion for the comparison of classifiers. Connections of our model with the theory of copulas are essential for this purpose. An application of the proposed method is developed in Section 6. Namely, we compare some classifiers for the detection of purchasers of home insurances. To conclude, a simulation study, which ‘corroborates’ the efficiency of model proposed in the manuscript, is developed in Section 7.

## 2 Notation and definitions

A stochastic order is a pre-order relation on a set of probabilities associated with a measurable space. Basically, a stochastic order is a criterion to rank probabilities. The reader is referred to the books Müller and Stoyan (2002),

Shaked and Shanthikumar (2007) and Belzunce *et al.* (2016), for an introduction to the theory of stochastic orders from theoretical and applied points of view.

Given a random vector or random variable  $W$ , its distribution function will be denoted by  $F_W$ . The survival function of  $W$  will be represented by  $\overline{F}_W$ . The expectation of  $W$  will be indicated by  $EW$ .

When  $W$  is a random variable, the symbol  $F_W^{-1}$  will stand for the quantile function of  $W$ , that is,  $F_W^{-1} : [0, 1) \rightarrow \overline{\mathbb{R}}$ , with  $F_W^{-1}(u) = \inf \{x \in \mathbb{R} : F_W(x) \geq u\}$  for any  $u \in [0, 1)$ . Note that  $F_W^{-1}(0) = -\infty$ . By agreement we define  $F_W(-\infty) = 0$ .

The dependence of the components of a random vector is given by the so-called copula. A copula can be defined as a distribution function of a random vector whose components follow uniform distribution on the interval  $(0, 1)$ .

If  $X = (X_1, X_2, \dots, X_n)$  is a random vector, then there exists a copula  $C$  such that  $F_X = C(F_{X_1}, F_{X_2}, \dots, F_{X_n})$ . We will denote by  $\widehat{C}$  the survival copula of  $C$ , that is, the copula associated with the survival function of  $X = (X_1, X_2, \dots, X_n)$ , namely,  $\overline{F}_{(X_1, X_2, \dots, X_n)} = \widehat{C}(\overline{F}_{X_1}, \overline{F}_{X_2}, \dots, \overline{F}_{X_n})$ .

The reader is referred, for instance, to Cherubini *et al.* (2004) and Nelsen (2006) for an introduction to the theory of copulas.

Some criteria to compare random vectors (stochastic orders) are included now. Let  $X = (X_1, X_2, \dots, X_n)$  and  $Y = (Y_1, Y_2, \dots, Y_n)$  be random vectors.

It is said that  $X$  is less than  $Y$  in the usual stochastic order, denoted by  $X \preceq_{st} Y$ , if  $E(f(X)) \leq E(f(Y))$  for all increasing mappings  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that the expectations exist, that is, for all mappings  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $f(x) \leq f(y)$  when  $x \leq y$ , where  $\leq$  stands for the usual componentwise order of  $\mathbb{R}^n$ .

We say that  $X$  is less than  $Y$  in the upper orthant order, denoted by  $X \preceq_{uo} Y$ , if  $\overline{F}_X(x_1, \dots, x_n) \leq \overline{F}_Y(x_1, \dots, x_n)$  for all  $(x_1, \dots, x_n) \in \mathbb{R}^n$ .

Let  $(X, Y)$  be a bidimensional random vector.

We say that  $X$  is positively regression dependent on  $Y$ , if  $P(X \leq t/Y = s)$  is decreasing in  $s$  for all  $t$ . Moreover,  $X$  is said to be right tail increasing in  $Y$ , if  $P(X > t/Y > s)$  is increasing in  $s$  for all  $t$  (see, for instance, Lehmann (1966) and Esary and Proschan (1972)).

The symbol  $\sim_{st}$  will mean the equality in distribution.

By  $\mathcal{B}(q)$  we will denote the Bernoulli distribution with parameter  $q \in (0, 1)$ . The uniform distribution on the interval  $(0, 1)$  will be indicated by

$U_{(0,1)}$ .

The support of a random variable will be denoted by *supp*.

Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be a distribution function,  $Im(F)$  will stand for the image set of  $F$ .

On the other hand, by ‘continuous random variable’ we will mean a random variable whose distribution function is continuous. Note that the existence of a density function is not required.

### 3 The mathematical model, preliminary results

In this section, we introduce the mathematical model to approach the motivating problem of this manuscript. That is based on what we have coined the alternative accumulated improvement curve. That curve tries to avoid the comparison of classifiers in subgroups which are different in size. Some preliminary results are also stated in this section.

**Definition 3.1.** *Let  $X$  be a random variable. Let  $p \in (0, 1]$ . We will denote by  $p^X$  the real value which satisfies that  $1 - p^X = F_X(F_X^{-1}(1 - p))$ .*

It is clear that  $p^X \leq p$  for any  $p \in (0, 1]$  since  $F_X(F_X^{-1}(1 - p)) \geq 1 - p$ . Moreover, if  $p_1 \leq p_2$  with  $p_1, p_2 \in (0, 1]$ , then  $p_1^X \leq p_2^X$  since  $F_X$  and  $F_X^{-1}$  are increasing. Observe that  $1 - p^X \in Im(F_X)$ .

It is not hard to prove that  $1 - p^X$  is the first value greater or equal to  $1 - p$  which belongs to  $Im(F_X)$ . As a consequence,  $p - p^X \leq P(F_X(X) = 1 - p^X)$  for any  $p \in (0, 1]$ .

We define the alternative accumulated improvement curve.

**Definition 3.2.** *Let  $X = (C_X, T_X)$  be a modelling vector. Let  $\widetilde{M}_X : (0, 1] \rightarrow \mathbb{R}$  be the mapping given by*

$$\begin{aligned} \widetilde{M}_X(p) = & \frac{1}{p} \left( p^{C_X} P(T_X = 1 / F_{C_X}(C_X) > 1 - p^{C_X}) \right. \\ & \left. + (p - p^{C_X}) P(T_X = 1 / F_{C_X}(C_X) = 1 - p^{C_X}) \right). \end{aligned}$$

*The mapping  $\widetilde{M}_X$  is said to be the alternative accumulated improvement curve of the modelling vector  $X$ .*

The motivation of this definition is the following. Given any  $p \in (0, 1]$ , we try to obtain the probability that the target assumes value 1 in the  $100p\%$  of the individuals with the largest values of the classifier  $C_X$ . If there exists exactly a group with such a percentage (in that case,  $p = p^{C_X}$ ), the above probability is calculated in those individuals. If such a subgroup does not exist (in that case,  $p \neq p^{C_X}$ ), on the one hand, it is taken the biggest group (with the largest values of the classifier) whose size is lower than  $100p\%$  (exactly  $100p^{C_X}\%$ ,  $(F_{C_X}(C_X) > 1 - p^{C_X})$ ), and on the other hand, that group is completed with the following group with the largest values of the classifier ( $F_{C_X}(C_X) = 1 - p^{C_X}$ ), taking the appropriate weighting.

A criterion to compare classifiers based on the alternative accumulated improvement curve is introduced now.

**Definition 3.3.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors. It will be said that  $X$  is less than  $Y$  in the alternative accumulated improvement curve stochastic order, if  $\widetilde{M}_X(p) \leq \widetilde{M}_Y(p)$  for any  $p \in (0, 1]$ . It will be denoted by  $X \preceq_{\widetilde{M}} Y$ .*

The relation  $X \preceq_{\widetilde{M}} Y$  means that for any  $p \in (0, 1]$ , given the  $100p\%$  of the individuals with the largest values of the corresponding classifiers  $C_X$  and  $C_Y$ , the probability of carrying out right classifications by estimating as 1 the value of the target at those individuals, is greater (at least the same) with classifier  $C_Y$  than with classifier  $C_X$ . Thus, modelling vector  $Y$  is better than  $X$  to classify individuals, whatever sizes of groups in which we compare those modelling vectors.

From now on,  $X \sim_{\widetilde{M}} Y$  will mean that  $X \preceq_{\widetilde{M}} Y$  and  $Y \preceq_{\widetilde{M}} X$  hold simultaneously.

Some technical results needed for our purposes are stated below. They relate values of  $p \in (0, 1]$  with the corresponding  $p^X$ . Their proofs can be found in the Appendix of the manuscript.

**Proposition 3.4.** *Let  $X$  be a random variable and  $p \in (0, 1]$ . Then,  $1 - p \in \text{Im}(F_X)$  if and only if  $p^X = p$ .*

**Proposition 3.5.** *Let  $X$  be a random variable and  $p \in (0, 1)$ . It holds that  $P(F_X(X) > 1 - p^X) = p^X$ .*

**Proposition 3.6.** *Let  $X$  be a random variable and  $p \in (0, 1]$ . It holds that  $p = p^X$  when  $P(F_X(X) = 1 - p^X) = 0$ .*

The following propositions are in relation to alternative accumulated improvement curves. Among other multiple purposes, they will be used to prove the continuity of alternative accumulated improvement curves, an interesting property from an applied point of view.

**Proposition 3.7.** *Let  $X = (C_X, T_X)$  be a modelling vector. Define the mapping  $H_X : (0, 1] \rightarrow \mathbb{R}$ , with  $H_X(p) = p - p\widetilde{M}_X(p)$  for any  $p \in (0, 1]$ . It holds that*

- i)  $H_X(p) = P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 0) + \alpha P(F_{C_X}(C_X) = 1 - p^{C_X}, T_X = 0)$ , where  $\alpha = \frac{p - p^{C_X}}{P(F_{C_X}(C_X) = 1 - p^{C_X})}$  if  $P(F_{C_X}(C_X) = 1 - p^{C_X}) > 0$ , otherwise  $\alpha = 0$ ,*
- ii)  $H_X(p) \leq P(T_X = 0)$  for any  $p \in (0, 1]$ ,*
- iii)  $H_X$  is increasing.*

*Proof.* In relation to *i)*, note that  $H_X(p) = p - p\widetilde{M}_X(p)$

$$\begin{aligned}
&= p - (p^{C_X} P(T_X = 1 / F_{C_X}(C_X) > 1 - p^{C_X}) \\
&\quad + (p - p^{C_X}) P(T_X = 1 / F_{C_X}(C_X) = 1 - p^{C_X})) \\
&= p^{C_X} (1 - P(T_X = 1 / F_{C_X}(C_X) > 1 - p^{C_X})) \\
&\quad + (p - p^{C_X}) (1 - P(T_X = 1 / F_{C_X}(C_X) = 1 - p^{C_X})) \\
&= p^{C_X} P(T_X = 0 / F_{C_X}(C_X) > 1 - p^{C_X}) \\
&\quad + (p - p^{C_X}) P(T_X = 0 / F_{C_X}(C_X) = 1 - p^{C_X}).
\end{aligned}$$

Proposition 3.5 reads that  $P(F_{C_X}(C_X) > 1 - p^{C_X}) = p^{C_X}$ , and thus  $H_X(p) = P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 0) + (p - p^{C_X}) P(T_X = 0 / F_{C_X}(C_X) = 1 - p^{C_X})$ .

This leads to *i)* by applying Proposition 3.6.

Condition *ii)* can be derived by *i)* and the fact that  $\alpha \in [0, 1]$ , observe that  $p - p^{C_X} \leq P(F_{C_X}(C_X) = 1 - p^{C_X})$ .

Regarding *iii)*, let  $p \leq p'$  with  $p, p' \in (0, 1]$ .

If  $p^{C_X} = p'^{C_X}$ , then  $H_X(p) \leq H_X(p')$  since  $p - p^{C_X} \leq p' - p^{C_X}$ .

If  $p^{C_X} < p'^{C_X}$ , we obtain that  $H_X(p) \leq P(F_{C_X}(C_X) \geq 1 - p^{C_X}, T_X = 0) \leq P(F_{C_X}(C_X) > 1 - p'^{C_X}, T_X = 0) \leq H_X(p')$ , which concludes the proof.  $\square$

**Proposition 3.8.** *Let  $X = (C_X, T_X)$  be a modelling vector. Define the mapping  $A_X : (0, 1] \rightarrow \mathbb{R}$ , with  $A_X(p) = p^{C_X}P(T_X = 1/F_{C_X}(C_X) > 1 - p^{C_X}) + (p - p^{C_X})P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X})$  for any  $p \in (0, 1]$ . Then,*

- i)  $A_X(p) = P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1) + (p - p^{C_X})P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X})$  for any  $p \in (0, 1]$ ,*
- ii)  $A_X$  is increasing,*
- iii)  $A_X(1) = q$ .*

*Proof.* If  $p^{C_X} = 0$ , the result is clear. If  $p^{C_X} \neq 0$ , by Proposition 3.5  $P(F_{C_X}(C_X) > 1 - p^{C_X}) = p^{C_X}$ , and thus,  $p^{C_X}P(T_X = 1/F_{C_X}(C_X) > 1 - p^{C_X}) = P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1)$ , which proves *i*).

Let us prove *ii*). Let  $p, p' \in (0, 1]$  with  $p \leq p'$ .

Firstly, let us suppose that  $p^{C_X} = p'^{C_X}$ . Applying formula *i*) of this proposition, we obtain that  $A_X(p) \leq A_X(p')$ .

Now, let  $p^{C_X} < p'^{C_X}$ . Observe that  $p - p^{C_X} \leq P(F_{C_X}(C_X) = 1 - p^{C_X})$ . Moreover, by Proposition 3.6, if  $P(F_{C_X}(C_X) = 1 - p^{C_X}) = 0$ , then  $p = p^{C_X}$ . Thus,  $A_X(p)$

$$\begin{aligned} &= P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1) + (p - p^{C_X})P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X}) \\ &\leq P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1) + P(F_{C_X}(C_X) = 1 - p^{C_X}, T_X = 1) \\ &= P(F_{C_X}(C_X) \geq 1 - p^{C_X}, T_X = 1) \\ &\leq P(F_{C_X}(C_X) > 1 - p'^{C_X}, T_X = 1) \leq A_X(p'). \end{aligned}$$

Therefore,  $A_X$  is increasing.

To conclude, note that for any random variable  $W$ , it holds that  $1^W = 1$ . Thus,  $A_X(1) = P(F_{C_X}(C_X) > 0, T_X = 1) = P(T_X = 1) = q$ .  $\square$

## 4 The continuity modelling vector

A key tool for the analysis and the development of the alternative accumulated improvement curve criterion, is constructed in this section. We will prove that given any modelling vector, there exists another modelling vector whose classifier follows uniform distribution on the interval  $(0, 1)$ , and both modelling vectors have the same alternative accumulated improvement curve, that is, they are ‘at the same level’ in the new criterion for the comparison of classifiers. That modelling vector has been coined as the continuity modelling vector of the former. This result and its proof are key to derive other

results. Note that the continuity of the classifier simplifies the formula of the alternative accumulated improvement curve since  $p^{C_X} = p$  for any  $p \in (0, 1]$  (see Proposition 3.4), and thus the analysis of the criterion.

**Theorem 4.1.** *Let  $X = (C_X, T_X)$  be a modelling vector with  $T_X \sim_{st} \mathcal{B}(q)$ . Then, there exists a modelling vector  $X' = (C'_X, T'_X)$  with  $T'_X \sim_{st} \mathcal{B}(q)$ , such that  $C'_X \sim_{st} U_{(0,1)}$  and  $X \sim_{\widetilde{M}} X'$ .*

*Proof.* Let us define  $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ , with

$$G(k_1, k_2) = \begin{cases} 0 & \text{if } k_1 \geq 1 \text{ or } k_2 \geq 1, \\ (1 - k_1)\widetilde{M}_X(1 - k_1) & \text{if } 0 \leq k_1 < 1 \text{ and } 0 \leq k_2 < 1, \\ 1 - k_1 & \text{if } 0 \leq k_1 < 1 \text{ and } k_2 < 0, \\ q & \text{if } k_1 < 0 \text{ and } 0 \leq k_2 < 1, \\ 1 & \text{if } k_1 < 0 \text{ and } k_2 < 0, \end{cases}$$

for any  $(k_1, k_2) \in \mathbb{R}^2$ .

We have that,

*i)*  $G$  is decreasing in  $k_1$  and  $k_2$ . Observe that  $(1 - k_1)\widetilde{M}_X(1 - k_1) = A_X(1 - k_1)$ , and  $A_X$  is positive, increasing and  $A_X(1) = q$  (see Proposition 3.8). Moreover,  $0 \leq \widetilde{M}_X \leq 1$ , and so,  $(1 - k_1)\widetilde{M}_X(1 - k_1) \leq 1 - k_1$  for any  $k_1 \in [0, 1)$ .

*ii)* It is clear that  $\lim_{k_1, k_2 \rightarrow -\infty} G(k_1, k_2) = 1$ .

*iii)*  $\lim_{k_1 \rightarrow \infty} G(k_1, k_2) = 0$  for any  $k_2 \in \mathbb{R}$ , and  $\lim_{k_2 \rightarrow \infty} G(k_1, k_2) = 0$  for any  $k_1 \in \mathbb{R}$ .

*iv)* As we show below, it holds that  $G(k_1, k_2) - G(k'_1, k_2) - G(k_1, k'_2) + G(k'_1, k'_2) \geq 0$  for any  $(k_1, k_2), (k'_1, k'_2) \in \mathbb{R}^2$  with  $k_1 \leq k'_1$  and  $k_2 \leq k'_2$ .

If  $k'_1 < 0$ , then  $G(k_1, k_2) = G(k'_1, k_2)$  and  $G(k_1, k'_2) = G(k'_1, k'_2)$ , and so, we have the result in this case.

The cases in which  $k'_2 < 0$ , or  $k'_1 \geq 1$ , or  $k'_2 \geq 1$ , or  $k_1 \geq 1$ , or  $k_2 \geq 1$  are also immediate.

The cases  $k_1 < 0$  and  $0 \leq k_1 < 1$  can be derived since the mapping  $H_X$  is increasing (see Proposition 3.7).

*v)* By the definition of  $G$ , it can be seen that  $\lim_{h \rightarrow 0^-} G(k_1 + h, k_2 + h) = G(k_1, k_2)$ .

Therefore,  $G$  is a survival function of a bidimensional random vector (see, for instance, Breiman (1992)). Let  $X' = (C'_X, T'_X)$  be a random vector whose survival function is  $G$ , that is,  $\overline{F}_{(C'_X, T'_X)} = G$ .

We have that

$$\overline{F}_{C'_X}(x) = \lim_{y \rightarrow -\infty} G(x, y) = \begin{cases} 0 & \text{if } x \geq 1, \\ 1 - x & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x < 0. \end{cases}$$

Therefore,  $C'_X$  follows uniform distribution on the interval  $(0, 1)$ . Moreover,

$$\overline{F}_{T'_X}(y) = \lim_{x \rightarrow -\infty} G(x, y) = \begin{cases} 0 & \text{if } y \geq 1, \\ q & \text{if } 0 \leq y < 1, \\ 1 & \text{if } y < 0, \end{cases}$$

that is,  $T'_X \sim_{st} \mathcal{B}(q)$ .

Thus,  $X' = (C'_X, T'_X)$  is a modelling vector with  $C'_X \sim_{st} U_{(0,1)}$  and  $T'_X \sim_{st} \mathcal{B}(q)$ .

To conclude, we need to prove that  $X \sim_{\widetilde{M}} X'$ . That is the same as  $\widetilde{M}_{X'} = \widetilde{M}_X$ . Let  $p \in (0, 1)$ . Since  $C'_X \sim_{st} U_{(0,1)}$ , we obtain that

$$\begin{aligned} \widetilde{M}_{X'}(p) &= P(T'_X = 1 / F_{C'_X}(C'_X) > 1 - p) = P(T'_X = 1 / C'_X > 1 - p) \\ &= \frac{P(C'_X > 1 - p, T'_X = 1)}{p} = \frac{1}{p} P(C'_X > 1 - p, T'_X > \frac{1}{2}) = \frac{1}{p} \overline{F}_{(C'_X, F'_X)}(1 - p, \frac{1}{2}) \\ &= \frac{1}{p} G(1 - p, \frac{1}{2}) = \frac{1}{p} (1 - (1 - p)) \widetilde{M}_X(1 - (1 - p)) = \widetilde{M}_X(p). \end{aligned}$$

When  $p = 1$ , note that  $\widetilde{M}_X(1) = A_X(1) = q$  for any modelling vector (see Proposition 3.8), which concludes the proof.  $\square$

**Definition 4.2.** Let  $X = (C_X, T_X)$  be a modelling vector with  $T_X \sim_{st} \mathcal{B}(q)$ . A modelling vector  $X' = (C'_X, T'_X)$  is said to be a continuity modelling vector of  $X = (C_X, T_X)$ , if  $T'_X \sim_{st} \mathcal{B}(q)$ ,  $C'_X \sim_{st} U_{(0,1)}$  and  $X \sim_{\widetilde{M}} X'$ .

The tool introduced in the above definition, and the particular continuity modelling vector constructed in Theorem 4.1, will be essential for different purposes of the manuscript.

As an example of application, we prove the continuity of alternative accumulated improvement curves immediately.

**Proposition 4.3.** Let  $X = (C_X, T_X)$  be a modelling vector. Then  $\widetilde{M}_X$  is continuous.

*Proof.* Let  $X' = (\widetilde{C'_X}, \widetilde{T'_X})$  be a continuity modelling vector of  $X = (C_X, T_X)$ . Recall that  $\widetilde{M_X} = \widetilde{M_{X'}}$ . Since  $C'_X$  is continuous, for any  $p \in (0, 1]$ ,  $\widetilde{M_{X'}}(p) = P(T_X = 1/F_{C'_X}(C'_X) > 1 - p) = \frac{1}{p}P(F_{C'_X}(C'_X) > 1 - p, T_X = 1)$ , which is continuous since  $F_{C'_X}(C'_X) \sim U_{(0,1)}$ .  $\square$

Another instance of applicability of the continuity modelling vector is the analysis of the preservation of the order under weak convergence. The following example, based on those modelling vectors, shows that the new criterion is not closed under that convergence. However, conditions under which the new stochastic order is closed under weak convergence, are stated below.

**Example 4.4.** Let  $X_n = (C_{X_n}, T_{X_n})$ ,  $n \in \mathbb{N}$ , be modelling vectors such that  $P(C_{X_n} = 1, T_{X_n} = 1) = 3/8$ ,  $P(C_{X_n} = 1, T_{X_n} = 0) = 1/8$ ,  $P(C_{X_n} = 1 - 1/n, T_{X_n} = 1) = 1/8$ , and  $P(C_{X_n} = 1 - 1/n, T_{X_n} = 0) = 3/8$ , for any  $n \in \mathbb{N} \setminus \{1\}$ . Observe that  $T_{X_n} \sim_{st} \mathcal{B}(1/2)$ .

It is not hard to see that for all  $n \in \mathbb{N}$  and  $p \in (0, 1]$ ,

$$\widetilde{M_{X_n}}(p) = \begin{cases} \frac{3}{4} & \text{if } 0 < p \leq 1/2, \\ \frac{1}{p}(\frac{3}{8} + (p - \frac{1}{2})\frac{1}{4}) & \text{if } 1/2 < p \leq 1. \end{cases}$$

In accordance with Theorem 4.1, there exists a common continuity modelling vector of all those vectors, say  $W = (C_W, T_W)$ , with  $C_W \sim_{st} U_{(0,1)}$ ,  $T_W \sim_{st} \mathcal{B}(1/2)$ , and  $X_n \sim_{\widetilde{M}} W$  for all  $n \in \mathbb{N}$ .

Let  $X = (C_X, T_X)$  be a modelling vector where  $C_X = 1$  *a.s.* and  $T_X \sim_{st} \mathcal{B}(1/2)$ . It is clear that  $\lim_n X_n = X$  in distribution.

However, the relation  $X \sim_{\widetilde{M}} W$  is false since  $\widetilde{M_X}(p) = 1/2$  for any  $p \in (0, 1]$ , observe that  $C_X$  and  $T_X$  are independent.

Therefore, the order  $\preceq_{\widetilde{M}}$  is not closed under weak convergence.

Conditions under which the new order is closed under weak convergence are stated below. The proof is included in the Appendix.

**Proposition 4.5.** *Let  $X_m = (C_{X_m}, T_{X_m})$ ,  $Y_m = (C_{Y_m}, T_{Y_m})$ ,  $m \in \mathbb{N}$ ,  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors. If  $X_m \preceq_{\widetilde{M}} Y_m$  for any  $m \in \mathbb{N}$ ,  $\lim_m X_m = X$  and  $\lim_m Y_m = Y$  in distribution, and  $C_X$  and  $C_Y$  are continuous, then  $X \preceq_{\widetilde{M}} Y$ .*

To conclude this section, next result shows another construction of continuity modelling vectors when classifiers are continuous.

**Proposition 4.6.** *Let  $X = (C_X, T_X)$  be a modelling vector, and let  $X' = (C'_X, T'_X)$  be a continuity modelling vector of  $X$ . If  $C_X$  is continuous, then  $(C'_X, T'_X) \sim_{st} (F_{C_X}(C_X), T_X)$ .*

*Proof.* Let  $Y = (F_{C_X}(C_X), T_X)$ . Since  $F_{C_X}(C_X) \sim_{st} U_{(0,1)}$ ,  $p^{F_{C_X}(C_X)} = p$  for any  $p \in (0, 1)$  (see Proposition 3.4), and so,  $\widetilde{M}_Y(p) = P(T_X = 1/F_{C_X}(C_X) > 1 - p) = \widetilde{M}_X(p) = \widetilde{M}_{X'}(p)$ .

Observe that for any  $k \in [0, 1)$  and  $p \in (0, 1)$ , it holds that  $P(F_{C_X}(C_X) > 1 - p, T_X > k) = P(F_{C_X}(C_X) > 1 - p)P(T_X > k/F_{C_X}(C_X) > 1 - p) = p\widetilde{M}_Y(p) = p\widetilde{M}_X(p)$ .

Now, it is immediately seen that the survival function of  $Y$ ,  $\overline{F}_Y$ , is equal to the mapping  $G$  in the proof of Theorem 4.1. Therefore,  $(C'_X, T'_X) \sim_{st} (F_{C_X}(C_X), T_X)$ .  $\square$

## 5 Characterizations of the new comparison criterion for classifiers and some consequences

Next, we prove different characterizations of the new criterion for the comparison of classification systems. Connections of that criterion with the theory of copulas will be essential for that purpose. We will also obtain sufficient conditions to assure the order between two classifiers, and some consequences.

The following result is key to state one of those characterizations.

**Theorem 5.1.** *Let  $X = (C_X, T_X)$  be a modelling vector, and let  $X' = (C'_X, T'_X)$  be a continuity modelling vector of  $X$ . Let  $C_{CMV}$  be a copula of  $X'$ . Then,  $C_{CMV}$  is a copula of  $X$ .*

*Proof.* Let  $C_{CMV}$  be a copula of  $X'$ , and  $\widehat{C}_{CMV}$  the corresponding survival copula. Let  $C$  be a copula of  $X$ , and  $\widehat{C}$  the associated survival copula.

Let us analyze the value of the survival distribution function of  $X$ .

Let  $y \in [0, 1)$  and  $x \in \mathbb{R}$  such that  $P(C_X > x) \in (0, 1)$ . We have that  $\overline{F}_X(x, y) = P(C_X > x, T_X > y) = \widehat{C}(\overline{F}_{C_X}(x), \overline{F}_{T_X}(y)) = \widehat{C}(p, q)$  with  $p = \overline{F}_{C_X}(x)$ . Then,  $1 - p = P(C_X \leq x)$ , and so  $p^{C_X} = p$  (see Proposition 3.4) and  $(F_{C_X}(C_X) > 1 - p) = (C_X > x)$  *a.s.*

Thus,  $\widetilde{M}_X(p) = P(T_X = 1/C_X > x) = \widetilde{M}_{X'}(p) = P(T'_X = 1/C'_X > 1 - p)$ . This implies that  $P(C_X > x, T_X = 1) = P(C'_X > 1 - p, T'_X = 1)$ .

Observe that  $P(C'_X > 1 - p, T'_X = 1) = \widehat{C}_{CMV}(\overline{F}_{C'_X}(1 - p), \overline{F}_{T'_X}(y)) = \widehat{C}_{CMV}(p, q)$ .

Therefore, we conclude that  $\overline{F}_X(x, y) = \widehat{C}_{CMV}(\overline{F}_{C_X}(x), \overline{F}_{T_X}(y))$  when  $x$  and  $y$  satisfy the above conditions. For other values of  $x$  and  $y$ , the same relation is trivial. Thus,  $\widehat{C}_{CMV}$  is a copula of  $X = (C_X, T_X)$ .  $\square$

The converse of Theorem 5.1 is not true, as we will see after Theorem 5.5. However, the converse holds if classifiers are continuous.

**Proposition 5.2.** *Let  $X = (C_X, T_X)$  be a modelling vector, and let  $X' = (C'_X, T'_X)$  be a continuity modelling vector of  $X$ . Let  $C$  be a copula of  $X$ . If  $C_X$  is continuous, then  $C$  is a copula of  $X'$ .*

*Proof.* Let  $C$  be a copula of  $X = (C_X, T_X)$ . Since  $F_{C_X}$  is increasing, Theorem 3.3 in Cai and Wei (2012) reads that  $C$  is also a copula of  $(F_{C_X}(C_X), T_X)$ . By Proposition 4.6,  $(C'_X, T'_X) \sim_{st} (F_{C_X}(C_X), T_X)$ , which proves the result.  $\square$

The following result states a characterization of the new criterion by means of survival copulas of continuity modelling vectors.

**Proposition 5.3.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors. Let  $X' = (C'_X, T'_X)$  and  $Y' = (C'_Y, T'_Y)$  be continuity modelling vectors of  $X$  and  $Y$ , respectively. Let  $\widehat{C}_{CMV}^X$  and  $\widehat{C}_{CMV}^Y$  be survival copulas of  $X'$  and  $Y'$ , respectively. Then,  $X \preceq_{\widetilde{M}} Y$  if and only if  $\widehat{C}_{CMV}^X(p, q) \leq \widehat{C}_{CMV}^Y(p, q)$  for any  $p \in [0, 1]$ .*

*Proof.* Since  $C'_X \sim_{st} C'_Y \sim_{st} U_{(0,1)}$ , we have that  $X' \preceq_{\widetilde{M}} Y'$  if and only if  $P(C'_X > 1 - p, T'_X = 1) \leq P(C'_Y > 1 - p, T'_Y = 1)$  for any  $p \in (0, 1)$ . This is the same as  $P(C'_X > 1 - p, T'_X > 0) \leq P(C'_Y > 1 - p, T'_Y > 0)$ , equivalently,  $\widehat{C}_{CMV}^X(\overline{F}_{C'_X}(1 - p), \overline{F}_{T'_X}(0)) \leq \widehat{C}_{CMV}^Y(\overline{F}_{C'_Y}(1 - p), \overline{F}_{T'_Y}(0))$  for any  $p \in (0, 1)$ . Note that  $\overline{F}_{C'_X}(1 - p) = \overline{F}_{C'_Y}(1 - p) = p$ , and  $\overline{F}_{T'_X}(0) = \overline{F}_{T'_Y}(0) = q$ . Therefore,  $X' \preceq_{\widetilde{M}} Y'$  if and only if  $\widehat{C}_{CMV}^X(p, q) \leq \widehat{C}_{CMV}^Y(p, q)$  for any  $p \in (0, 1)$ . Observe that if  $p = 0$  or  $p = 1$ , both expressions are the same. Now, the result follows from the definition of continuity modelling vector.  $\square$

We relate the new method for rating classifiers with that proposed in López-Díaz *et al.* (2017). That will lead to other characterizations of  $\preceq_{\widetilde{M}}$ .

**Proposition 5.4.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors such that  $C_X$  and  $C_Y$  are continuous. Then,  $X \preceq_{\widetilde{M}} Y$  if and only if  $X \preceq_M Y$ .*

*Proof.* If  $C_X$  is continuous,  $\widetilde{M}_X(p) = P(T_X = 1/F_{C_X}(C_X) > 1-p) = P(T_X = 1/F_{C_X}(C_X) \geq 1-p) = M_X(p)$  for any  $p \in (0, 1)$ , which proves the result.  $\square$

It is not hard to see that both criteria are not the same.

Now, we are in conditions to state the following characterizations of the alternative accumulated improvement curve criterion.

**Theorem 5.5.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors. Let  $X' = (C'_X, T'_X)$  and  $Y' = (C'_Y, T'_Y)$  be continuity modelling vectors of  $X$  and  $Y$ , respectively. Let  $\widehat{C}_{CMV}^X$  and  $\widehat{C}_{CMV}^Y$  be survival copula functions of  $X'$  and  $Y'$ , respectively. The following conditions are equivalent,*

- i)  $X \preceq_{\widetilde{M}} Y$ ,*
- ii)  $X' \preceq_{\widetilde{M}} Y'$ ,*
- iii)  $\widehat{C}_{CMV}^X(p, q) \leq \widehat{C}_{CMV}^Y(p, q)$  for any  $p \in [0, 1]$ ,*
- iv)  $X' \preceq_{uo} Y'$ ,*
- v)  $T'_X | (C'_X \geq 1-p) \preceq_{st} T'_Y | (C'_Y \geq 1-p)$  for any  $p \in (0, 1)$ ,*
- vi) for any  $x \in \text{supp}(T'_X | C'_X \geq 1-p) \cap \text{supp}(T'_Y | C'_Y \geq 1-p)$  and  $p \in (0, 1)$ ,  
 $F_{T'_X | C'_X \geq 1-p}^{-1}(F_{T'_Y | C'_Y \geq 1-p}(x)) \leq x$ .*

*Proof.* Conditions *i)* and *ii)* are equivalent by Definition 4.2.

The equivalence of *i)* and *iii)* follows from Proposition 5.3.

The equivalence of *ii)* and *iv)* is because of Proposition 5.4, and Proposition 3.15 in López-Díaz *et al.* (2017), note that the classifiers of  $X'$  and  $Y'$  have uniform distribution on the interval  $(0, 1)$ . In the same way, the equivalence of *ii)* and *v)*, and the equivalence of *ii)* and *vi)* are due to Propositions 3.14 and 3.13, respectively, in López-Díaz *et al.* (2017).  $\square$

The following characterizations hold when classifiers are continuous.

**Theorem 5.6.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors. Let  $C_1$  and  $C_2$  be copula functions of  $X$  and  $Y$ , respectively. If  $C_X$  and  $C_Y$  are continuous, the following conditions are equivalent,*

- i)  $X \preceq_{\widetilde{M}} Y$ ,*
- ii)  $C_1(p, 1-q) \leq C_2(p, 1-q)$  for any  $p \in [0, 1]$ ,*

iii)  $\widehat{C}_1(p, q) \leq \widehat{C}_2(p, q)$  for any  $p \in [0, 1]$ .

*Proof.* Let  $X' = (C'_X, T'_X)$  and  $Y' = (C'_Y, T'_Y)$  be continuity modelling vectors of  $X$  and  $Y$ , respectively. By Proposition 5.2,  $C_1$  is a copula of  $X'$  and  $C_2$  of  $Y'$ . This proves the equivalence between *i*) and *iii*) using Theorem 5.5. The equivalence between *ii*) and *iii*) follows from the relation between a copula and its survival copula.  $\square$

Let us see that the converse of Theorem 5.1 is not true. Suppose that the converse holds. Let  $C$  be a copula of a modelling vector  $X = (C_X, T_X)$ . Let us consider the mapping  $h : \mathbb{R} \rightarrow \mathbb{R}$ , with  $h = 0$ . Since  $h$  is increasing, Theorem 3.3 in Cai and Wei (2012) says that  $C$  is a copula of the modelling vector  $(0, T_X)$ . Under our assumption,  $C$  is a copula of  $(0', T'_X)$ , a continuity modelling vector of  $(0, T_X)$ , and a copula of  $X' = (C'_X, T'_X)$ . Applying Theorem 5.5, we obtain that  $(C_X, T_X) \sim_{\widetilde{M}} (0, T_X)$ , which obviously is not true in general since 0 and  $T_X$  are independent.

The following result states sufficient conditions for the new criterion to compare classifiers.

**Proposition 5.7.** *Let  $X = (C_X, T_X)$  and  $Y = (C_Y, T_Y)$  be modelling vectors such that  $C_Y$  is continuous. If  $(F_{C_X}(C_X), T_X) \preceq_{uo} (F_{C_Y}(C_Y), T_Y)$ , then  $X \preceq_{\widetilde{M}} Y$ .*

*Proof.* Let us see that  $\widetilde{M}_X(p) \leq \widetilde{M}_Y(p)$  for any  $p \in (0, 1]$ . This is the same as  $A_X(p) \leq A_Y(p)$  for any  $p \in (0, 1)$  (see Proposition 3.8).

Observe that by Proposition 3.5,  $A_X(p)$

$$= P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1) + (p - p^{C_X})P(T_X = 1 / F_{C_X}(C_X) = 1 - p^{C_X}).$$

If  $p = p^{C_X}$ , then  $A_X(p) = P(F_{C_X}(C_X) > 1 - p, T_X = 1)$ . If  $p \neq p^{C_X}$ , recall that  $p - p^{C_X} \leq P(F_{C_X}(C_X) = 1 - p^{C_X})$ , and so, it holds that  $A_X(p) \leq P(F_{C_X}(C_X) \geq 1 - p^{C_X}, T_X = 1) \leq P(F_{C_X}(C_X) > 1 - p, T_X = 1)$ .

Since  $C_Y$  is continuous,  $1 - p \in \text{Im}(F_{C_Y})$  for any  $p \in (0, 1)$ , thus,  $p^{C_Y} = p$  (see Proposition 3.4) and  $F_{C_Y}(C_Y) \sim_{st} U_{(0,1)}$ . Then,  $A_Y(p) = pP(T_Y = 1 / F_{C_Y}(C_Y) > 1 - p) = P(F_{C_Y}(C_Y) > 1 - p, T_Y = 1)$ . Now, the result follows from the condition  $(F_{C_X}(C_X), T_X) \preceq_{uo} (F_{C_Y}(C_Y), T_Y)$ .  $\square$

An interesting property that an alternative accumulated improvement curve can satisfy, is being decreasing. Roughly speaking, if such a curve decreases, taking smaller groups of the population with larger values of the

classifier, increases the probability that the target assumes value 1. That means that the probability of being right when we estimate the value of the target as 1, increases as  $p$  approaches to 0. In the following results, we state conditions on modelling vectors to have decreasing alternative accumulated improvement curves. The proofs of those results are included in the Appendix.

**Proposition 5.8.** *Let  $X = (C_X, T_X)$  be a modelling vector. Let  $T_X$  be positively regression dependent on  $C_X$ . Then,*

- i) *the mapping  $\widetilde{M}_X$  is decreasing,*
- ii)  *$\widetilde{X} \preceq_{\widetilde{M}} X$  where  $\widetilde{X} = (\widetilde{C}_X, \widetilde{T}_X)$ , with  $\widetilde{C}_X \sim_{st} C_X$ ,  $\widetilde{T}_X \sim_{st} T_X$ ,  $\widetilde{C}_X$  and  $\widetilde{T}_X$  being independent.*

**Proposition 5.9.** *Let  $X = (C_X, T_X)$  be a modelling vector where  $C_X$  is a continuous random variable. Let  $C$  be a copula of  $X$ . Let us define the mapping  $g : (0, 1) \rightarrow \mathbb{R}$  with  $g(x) = \widehat{C}(x, q)/x$ , where  $\widehat{C}$  stands for the survival copula of  $C$ . If  $g$  is decreasing, so is  $\widetilde{M}_X$ .*

Copulas which satisfy the condition of  $g$  in Proposition 5.9 are, for instance,  $C : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ , with  $\widehat{C}(u, v) = uv$ ,  $\widehat{C}(u, v) = \min\{u, v\}$ , or those whose survival copulas belong to the Cuadras-Augé family. That family is given by  $C_\theta : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ , with

$$C_\theta(u, v) = (\min\{u, v\})^\theta (uv)^{1-\theta} = \begin{cases} uv^{1-\theta} & \text{if } u \leq v, \\ u^{1-\theta}v & \text{if } v \leq u, \end{cases}$$

for any  $u, v \in [0, 1]$ , with  $\theta \in [0, 1]$  (see Cuadras and Augé (1981)).

It is worth noting that the characterizations of the criterion  $\preceq_{\widetilde{M}}$  permit to find families of modelling vectors which are a lattice in  $\preceq_{\widetilde{M}}$ . Let us see an example of that, its proof is in the Appendix.

**Proposition 5.10.** *Let  $\Upsilon$  be the set of modelling vectors whose targets follow Bernoulli distribution with parameter  $q$ , classifiers are continuous, and have a survival copula in the Cuadras-Augé family. Then,  $(\Upsilon, \preceq_{\widetilde{M}})$  is a lattice.*

The following propositions show maximal and minimal modelling vectors in the set of modelling vectors whose targets follow Bernoulli distribution with parameter  $q$ . As expected, the target and one minus the target are the best and worst classifiers, respectively. Proofs are included in the Appendix.

**Proposition 5.11.** *Let  $X = (C_X, T_X)$  be a modelling vector with  $T_X \sim_{st} \mathcal{B}(q)$ . Let  $Z$  be the modelling vector  $(T_Z, T_Z)$  with  $T_Z \sim_{st} \mathcal{B}(q)$ . Then,  $X \preceq_{\widetilde{M}} Z$ .*

**Proposition 5.12.** *Let  $X = (C_X, T_X)$  be a modelling vector with  $T_X \sim_{st} \mathcal{B}(q)$ . Let  $W$  be the modelling vector  $(1 - T_W, T_W)$  with  $T_W \sim_{st} \mathcal{B}(q)$ . Then,  $W \preceq_{\widetilde{M}} X$ .*

To conclude this section, we compare the alternative accumulated improvement curve criterion with one of the most popular rating systems, that based on the CAP curve of a classifier.

Let  $(C, T)$  be a modelling vector where  $C$  can take only the values 0 and 1, and the value of  $T$  is estimated by the value of  $C$ . The sensitivity of  $(C, T)$  is the ratio between the so-called true positives and positive condition, that is,  $P(C = 1, T = 1)/P(T = 1)$ . Given  $(C, T)$  a modelling vector and  $k \in \mathbb{R}$ , let  $(C_k, T)$  be the modelling vector given by  $C_k = 1$  when  $C \geq k$ , otherwise  $C_k = 0$ . For each  $(C_k, T)$ , consider its sensitivity, denoted by  $Sens(k)$ . The CAP curve of  $(C, T)$  is the set of points  $\{(P(C \geq k), Sens(k)) \mid k \in \mathbb{R}\}$ . The CAP index is the area under the CAP curve.

In López-Díaz *et al.* (2017), it is proved that when the classifier of a modelling vector  $(C, T)$  is continuous, then

$$Sens(k) = M_{(C,T)}(p_k) \frac{p_k}{q},$$

where  $P(C \geq k) = p_k$  with  $p_k \in (0, 1)$  (see the proof of Proposition 3.30 in the above reference). Thus, the CAP curve of  $(C, T)$  is  $\{(p_k, M_{(C,T)}(p_k) \frac{p_k}{q}) \mid k \in \mathbb{R}\}$ .

By Proposition 5.4, when  $C$  is continuous, the CAP curve of  $(C, T)$  is  $\{(p_k, \widetilde{M}_{(C,T)}(p_k) \frac{p_k}{q}) \mid k \in \mathbb{R}\}$ .

Assume now that  $(C, T)$  is any modelling vector. Let  $(C', T')$  be a continuity modelling vector of  $(C, T)$ . Recall that  $C' \sim_{st} U_{(0,1)}$ . Thus, the CAP curve of  $(C', T')$  is  $\{(p, \widetilde{M}_{(C',T')}(p) \frac{p}{q}) \mid p \in (0, 1)\} = \{(p, \widetilde{M}_{(C,T)}(p) \frac{p}{q}) \mid p \in (0, 1)\}$ , since  $\widetilde{M}_{(C',T')} = \widetilde{M}_{(C,T)}$ .

A natural question appears now. Let  $H_C : (0, 1) \rightarrow \mathbb{R}$ , with  $H_C(p) = \widetilde{M}_{(C,T)}(p) \frac{p}{q}$  for any  $p \in (0, 1)$ . Which is the relation between the CAP curve of  $(C, T)$  and  $H_C$ ?, that is, which is the connection between the CAP curve of  $(C, T)$  and the CAP curve of any continuity modelling vector of  $(C, T)$ ? The following result answers this question. Its proof is in the Appendix.

**Proposition 5.13.** *Let  $(C, T)$  be a modelling vector and let  $p \in (0, 1)$ .*

- i) If there exists  $k \in \mathbb{R}$  such that  $P(C \geq k) = p$ , the value of the CAP curve of  $(C, T)$  at  $k$  satisfies that  $(P(C \geq k), \text{Sens}(k)) = (p, H_C(p))$ ,*
- ii) if there is not  $k \in \mathbb{R}$  such that  $P(C \geq k) = p$ ,  $H_C(p)$  is the value at  $p$  of the linear interpolation constructed with the points  $(p^C, H_C(p^C))$  and  $(p^C + m, H_C(p^C + m))$ , where  $m = P(F_C(C) = 1 - p^C)$ .*

## 6 Application to the comparison of classifiers for the sale of home insurances

The model introduced in this manuscript, has been applied to the comparison of the performance of some classifiers for the estimation of possible purchasers of a home insurance in a banking institution. For that analysis, we have made use of a database provided by Liberbank (a Spanish commercial banking).

In the study, we considered clients of that banking that, at 31 December 2016, are aged between 18 and 80 years, have at least 500 euros of business balance (savings+debts), an account balance greater than 0, and do not enjoy of a home insurance (usual legal commercial filters were applied). In total, 740.315 clients were included. The number of available variables of each individual was 58.

We aim to estimate the target  $T =$  ‘the client will buy a home insurance in 2017’. It was considered that a client had bought a home insurance if at 31 December 2017, he had one. The proportion of purchasers of a home insurance in the sample was 0.006.

The database was divided at random into two groups, a training group containing 540.315 clients, and a testing or validation group with 5.000.000 of observations taking at random from the remaining group of 200.000 individuals. The training group was used for the construction of the classifiers, the testing group for the comparison of the performance of those classifiers.

We have compared eight classifiers constructed by means of the available variables. Some of them were designed using three important variables in accordance with the opinion of some experts of the bank, namely, assets, number of bank products and age.

The scheme to construct classifiers is as follows. Let  $(X_1, X_2, \dots, X_{58})$  be the random vector associated with the 58 available variables of each individual, that is,  $X_i$  stands for the random variable associated with the  $i^{th}$

characteristic. By means of the above random vector, we construct eight different classifiers  $C_i$  as

$$C_i = H_i(X_1, X_2, \dots, X_{58}), \quad 1 \leq i \leq 8.$$

Each  $H_i$  is developed by means of decision trees or logistic regression (machine learning techniques).

The classifiers are:

- Classifier 1 ( $C_1$ ): decision tree with all the variables, 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients.
- Classifier 2 ( $C_2$ ): logistic regression with all the variables and stepwise selection method.
- Classifier 3 ( $C_3$ ): decision tree with the variables assets, number of bank products and age, 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients.
- Classifier 4 ( $C_4$ ): logistic regression with the variables assets, number of bank products and age, and stepwise selection method.
- Classifier 5 ( $C_5$ ): decision tree with the variable age, 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients.
- Classifier 6 ( $C_6$ ): decision tree with the variable age, 2 branches per division, maximum depth of 3 and minimum size of leaf equals to 5.000 clients.
- Classifier 7 ( $C_7$ ): decision tree with the variable assets, 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients.
- Classifier 8 ( $C_8$ ): decision tree with the variable assets, 2 branches per division, maximum depth of 3 and minimum size of leaf equals to 5.000 clients.

For the comparison of any two classifiers  $C_i$  and  $C_j$ , the testing group was divided at random one hundred times into two groups of size 2.500.000. We follow this procedure to avoid the influence of a particular division in the final conclusions.

In each division of the testing group, we proceeded as follows. Each part of the division was assigned at random to one of the classifiers. The sample alternative accumulated improvement curve of each classifier was calculated in a mesh of twenty points (20 values of  $p$ ), from 0.05 to 1, with a step of 0.05, using the corresponding part of the division. For each classifier  $C$  and each of the values of  $p$ , the sample alternative accumulated improvement curve was obtained as follows.

Let  $Z$  be a random variable with uniform distribution on the interval  $(0, 1)$ ,  $Z \sim_{st} U(0, 1)$ , independent of the modelling vector  $(C, T)$ .

Let  $k_p = 1 - \frac{p-p^C}{P(F_C(C)=1-p^C)}$  if  $P(F_C(C) = 1 - p^C) \neq 0$ , otherwise  $k_p = 1$ .  
Let

$$A^p = \{F_C(C) > 1 - p^C\} \cup \{F_C(C) = 1 - p^C, Z > k_p\}.$$

By Proposition 3.5 and 3.6, it is immediately seen that  $P(A^p) = p$ .

Let us see that  $\widetilde{M}_{(C,T)}(p) = P(T = 1/A^p)$ . Observe that

$$P(T = 1/A^p) = \frac{P(A^p, T = 1)}{P(A^p)}$$

$$= \frac{1}{p}(P(F_C(C) > 1 - p^C, T = 1) + P(F_C(C) = 1 - p^C, Z > k_p, T = 1)).$$

By Proposition 3.5 and 3.6, that is equal to  $P(T = 1/F_C(C) > 1 - p)$  if  $P(F_C(C) = 1 - p^C) = 0$  since  $p = p^C$ , and so that value is  $\widetilde{M}_{(C,T)}(p)$ . When  $P(F_C(C) \neq 1 - p^C) \neq 0$ , such a value is equal to

$$\frac{1}{p}(p^C P(T = 1/F_C(C) > 1 - p^C)$$

$$+ P(T = 1/F_C(C) = 1 - p^C)P(F_C(C) = 1 - p^C)P(Z > 1 - \frac{p - p^C}{P(F_C(C) = 1 - p^C)}),$$

that is,

$$\frac{1}{p}(p^C P(T = 1/F_C(C) > 1 - p^C) + (p - p^C)P(T = 1/F_C(C) = 1 - p^C)),$$

which is  $\widetilde{M}_{(C,T)}(p)$ .

This result (*i.e.*  $\widetilde{M}_{(C,T)}(p) = P(T = 1/A^p)$ ) is key to compute the sample alternative accumulated curve of any classifier  $C$  at any point  $p$ , since it means that  $\widetilde{M}_{(C,T)}(p)$  is a population proportion. Roughly speaking, the sample version of  $\widetilde{M}_{(C,T)}(p)$  is the proportion of ones of the target  $T$  in the sample counterpart of  $A^p = \{F_C(C) > 1 - p^C\} \cup \{F_C(C) = 1 - p^C, Z > k_p\}$ . Note that such a sample set is easy to calculate for a sample drawn from  $(C, T)$ . Basically, for a value  $p$ , we should obtain the sample value of  $p^C$  using the empirical distribution function of the sample of  $C$  (note that one minus such a value is the first value greater or equal to  $1 - p$  which belongs to the image of the empirical distribution function).

The sample version of  $A^p$  is given by those individuals in which the value of the classifier at the empirical distribution function of  $C$  is greater than one minus the sample version of  $p^C$  (this corresponds to  $(F_C(C) > 1 - p^C)$ ), and for those individuals in which the equality holds (that corresponds to  $(F_C(C) = 1 - p^C)$ ), they are included or not using the value of the random variable  $Z$  ( $Z > k_p$  or not).

This approach has a double advantage. On the one side, it permits to calculate the sample value of  $\widetilde{M}_{(C,T)}(p)$  easily. On the other hand (and extremely important for our purposes), it permits the inferential comparison of the values  $\widetilde{M}_{(C_i,T)}(p)$  and  $\widetilde{M}_{(C_j,T)}(p)$  for any two classifiers  $C_i$  and  $C_j$ , since that inferential procedure is reduced to a simple comparison of two proportions (note that  $\widetilde{M}_{(C,T)}(p)$  is the probability that a Bernoulli variable assumes value 1 in the subgroup of the population  $A^p$ ).

Thus, for each division of the testing group, the sample alternative accumulated improvement curves  $\widetilde{M}_{(C_i,T)}$  and  $\widetilde{M}_{(C_j,T)}$  were obtained at the corresponding values of  $p$  with the above procedure, that is, by means of the proportions of purchasers of a home insurance in the corresponding subgroups. Therefore, for the comparison of two classifiers  $C_i$  and  $C_j$ , we have 100 pairs of sample values of  $\widetilde{M}_{(C_j,T)}$  and  $\widetilde{M}_{(C_i,T)}$ .

We aim to analyze:

- if  $C_i$  and  $C_j$  are equally efficient to detect purchasers of home insurances, equivalently,  $(C_i, T) \sim_{\widetilde{M}} (C_j, T)$ , that is,  $\widetilde{M}_{(C_i,T)} = \widetilde{M}_{(C_j,T)}$ ,
- if  $C_i$  is more efficient than  $C_j$  to predict purchasers of home insurances, this is the same as  $(C_j, T) \preceq_{\widetilde{M}} (C_i, T)$ , or  $\widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$  (in the event of rejection of the same efficiency).

In relation to the same efficiency problem, for each division of the testing

group, we considered tests for the equality of the alternative accumulated improvement curves of both classifiers at the values of  $p$ . The null hypothesis of those tests is  $H_0 : \widetilde{M}_{(C_i,T)}(p) = \widetilde{M}_{(C_j,T)}(p)$ . Note that this is a test for the equality of two independent proportions.

We obtained 20  $p$ -values associated with the 20 values of  $p$  (from 0.05 to 1 with a step of 0.05). As a summary  $p$ -value of the hypothesis  $\widetilde{M}_{(C_i,T)} = \widetilde{M}_{(C_j,T)}$ , we took the smallest one among the 20  $p$ -values, that is, the  $p$ -value showing more evidence that the relation  $\widetilde{M}_{(C_i,T)} = \widetilde{M}_{(C_j,T)}$  is false. As usual, a Bonferroni correction to reduce the chances of obtaining false-positive results was considered. Taking the usual level of significance  $\alpha = 0.05$ ,  $p$ -values should be compared with  $0.05/20$ .

Therefore, we obtained 100  $p$ -values, one for each division of the testing group. As a representative  $p$ -value for the null hypothesis  $\widetilde{M}_{(C_i,T)} = \widetilde{M}_{(C_j,T)}$ , we considered the median  $p$ -value.

Table 1 shows our inferential conclusions. For ease of reading, it contains representative  $p$ -values multiplied by 20 (to be compared with 0.05 because of Bonferroni correction). For each row ( $C_i$ ) and each column ( $C_j$ ), the table displays  $p\text{-values} \times 20$  for the null hypothesis  $\widetilde{M}_{(C_j,T)} = \widetilde{M}_{(C_i,T)}$ .

We conclude that classifiers  $C_1$  and  $C_3$  are equally useful to predict purchasers of a home insurance. Classifiers  $C_7$  and  $C_8$  have also the same performance. On the other hand, there are not two other classifiers at the same level in the alternative accumulated improvement curve stochastic order.

In relation to the more efficient classifier issue, that is, the tests with null hypothesis  $H_0 : \widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$ , we followed the same steps. In this case, we used a test for the comparison  $\leq$  of two independent proportions at each value of  $p$ . Table 2 contains our results. For each row ( $C_i$ ) and each column ( $C_j$ ), we have the representative  $p$ -value multiplied by 20 (Bonferroni correction) for the null hypothesis  $\widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$ .

Inferential results can be summarize as follows:

- $(C_6, T) \preceq_{\widetilde{M}} (C_5, T) \preceq_{\widetilde{M}} (C_4, T) \preceq_{\widetilde{M}} (C_2, T) \preceq_{\widetilde{M}} (C_3, T) \sim_{\widetilde{M}} (C_1, T)$
- $(C_6, T) \preceq_{\widetilde{M}} (C_5, T) \preceq_{\widetilde{M}} (C_8, T) \sim_{\widetilde{M}} (C_7, T) \preceq_{\widetilde{M}} (C_3, T) \sim_{\widetilde{M}} (C_1, T)$

Inferential conclusions say that  $C_1$  and  $C_3$  are the best classifiers to predict purchasers of a home insurance and they are equally useful.

Classifier  $C_6$  is worse than the rest of classifiers. Classifier  $C_5$  is better than classifier  $C_6$ , but worse than the remaining classifiers.  $C_4$  is better than

$C_5$  and worse than  $C_2$  and  $C_1$  and  $C_3$ . Classifier  $C_2$  is worse than  $C_1$  and  $C_3$ . Finally,  $C_7$  and  $C_8$  are similar for our purposes, worse than  $C_1$  and  $C_3$  and better than  $C_5$ .

The diagram in Figure 1 represents the inferential conclusions. An arrow from  $(C_j, T)$  to  $(C_i, T)$  means that  $C_i$  shows a better performance than  $C_j$  to predict target  $T$ . Arrows which correspond to relations which can be obtained from the diagram applying transitivity, are omitted.

It is very interesting to remark that  $C_1$  and  $C_3$  are the best classifiers to predict purchasers of a home insurance, being equally effective. Recall that they are constructed with the same pattern, namely, decision tree, 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients. However,  $C_1$  uses all the variable (58), and  $C_3$  only three variables (assets, number of bank products and age). Obviously, from an applied point of view, we should use classifier  $C_3$  since its computing cost is much smaller than the cost of classifier  $C_1$ . On the other hand, the inferential results reinforce the opinion of the experts of the bank, which believe that assets, number of bank products and age, are relevant variables to predict purchasers of a home insurance.

The graphical representation of the sample alternative accumulated improvement curves of the eight classifiers appears in Figure 2. Those sample mappings were obtained with the whole testing group. Such mappings were calculated at the corresponding 20 values of  $p$ . They were depicted by means of interpolation. Observe that the curves of classifiers  $C_1$  and  $C_3$  are above the remaining curves in most points. Of course, that representation only permits to envisage possible relations among classifiers, and an inferential procedure (as the one previously developed) is necessary to draw conclusions.

We would like to emphasize that the proposed method in the present manuscript has been used by Liberbank to compare the performance of different classifiers.

From a computational point of view, all the calculations were carried out with the software SAS (SAS Base, SAS Stat and SAS Enterpriser Miner).

## 7 Simulation study

In this section, we develop a simulation study to ‘validate’ the procedure proposed in the manuscript to compare classifiers. That analysis is as follows.

Given a random variable  $W$ , let  $Q_\alpha^W$  stand for the quantile of order  $\alpha$  of

the variable  $W$ , with  $\alpha \in (0, 1)$ .

Let  $(C_1, T)$  be a modelling vector such that the conditional distribution of  $T$  given that  $C_1 = c$  is Bernoulli with parameter  $\exp(c)/(1 + \exp(c))$  (inverse logit transformation of  $c$ ).

Consider the following classifiers for the estimation of the target  $T$ :

- Classifier 1 ( $C_1$ ):  $C_1$  follows uniform distribution on the interval  $(-4, 0)$ ,
- Classifier 2 ( $C_2$ ): with  $C_2 = -4I_{(-\infty, Q_{0.5}^{C_1})}(C_1) + Q_{0.5}^{C_1}I_{[Q_{0.5}^{C_1}, +\infty)}(C_1)$ ,
- Classifier 3 ( $C_3$ ): with  $C_3 = -4I_{(-\infty, Q_{0.25}^{C_1})}(C_1) + Q_{0.25}^{C_1}I_{[Q_{0.25}^{C_1}, Q_{0.5}^{C_1})}(C_1) + Q_{0.5}^{C_1}I_{[Q_{0.5}^{C_1}, Q_{0.75}^{C_1})}(C_1) + Q_{0.75}^{C_1}I_{[Q_{0.75}^{C_1}, +\infty)}(C_1)$ ,
- Classifier 4 ( $C_4$ ): where  $C_4$  is given by  $C_4 = -4I_{(-\infty, Q_{0.125}^{C_1})}(C_1) + Q_{0.125}^{C_1}I_{[Q_{0.125}^{C_1}, Q_{0.25}^{C_1})}(C_1) + Q_{0.25}^{C_1}I_{[Q_{0.25}^{C_1}, Q_{0.375}^{C_1})}(C_1) + Q_{0.375}^{C_1}I_{[Q_{0.375}^{C_1}, Q_{0.5}^{C_1})}(C_1) + Q_{0.5}^{C_1}I_{[Q_{0.5}^{C_1}, Q_{0.625}^{C_1})}(C_1) + Q_{0.625}^{C_1}I_{[Q_{0.625}^{C_1}, Q_{0.75}^{C_1})}(C_1) + Q_{0.75}^{C_1}I_{[Q_{0.75}^{C_1}, Q_{0.875}^{C_1})}(C_1) + Q_{0.875}^{C_1}I_{[Q_{0.875}^{C_1}, +\infty)}(C_1)$ .

Note that  $C_2$ ,  $C_3$  and  $C_4$  are stepwise functions of  $C_1$ , where the steps are constructed by means of different quantiles of  $C_1$ . In the case of  $C_2$ , using the median of  $C_1$ ,  $C_3$  by means of the quantiles of order 0.25, 0.5 and 0.75, and with the quantiles of order 0.125, 0.25, 0.375, 0.5, 0.625, 0.75 and 0.875 in the case of  $C_4$ .

Roughly speaking,  $C_4$  is constructed by means of  $C_1$  with a loss of information,  $C_3$  is designed using  $C_1$  with a loss of information ‘greater’ than that of  $C_4$  since we use an smaller number of quantiles of  $C_1$ . In the case of  $C_2$ , the loss of information is even ‘greater’ since it uses only one quantile, the median of  $C_1$ .

As a consequence, any coherent rating system should provide the following ordering:  $C_1$  is better than  $C_4$  (not worse),  $C_4$  better than  $C_3$ , and this should be better than  $C_2$ . With the alternative accumulated improvement curve stochastic order, we should obtain the relation

$$(C_2, T) \preceq_{\widetilde{M}} (C_3, T) \preceq_{\widetilde{M}} (C_4, T) \preceq_{\widetilde{M}} (C_1, T).$$

We aim to analyze if such an assumption can be ‘validated’ by means of simulation.

For such a purpose, we have simulated a database with one million of observations of  $(C_1, T)$  and thus, a million of observations of  $(C_2, T)$ ,  $(C_3, T)$  and  $(C_4, T)$ .

Applying the procedure explained in Section 6, we have obtained the sample version of  $\widetilde{M}_{(C_1,T)}$ ,  $\widetilde{M}_{(C_2,T)}$ ,  $\widetilde{M}_{(C_3,T)}$  and  $\widetilde{M}_{(C_4,T)}$ . Recall that those sample mappings are sample proportions.

Those sample alternative accumulated improvement curves were calculated in a mesh of twenty points (20 values of  $p$ ), from 0.05 to 1, with a step of 0.05, using the whole set of simulations.

The graphical representation appears in Figure 3. That picture seems to corroborate the above relations among classifiers.

To derive a conclusion, we apply the inferential procedure described in Section 6. Note that in this case, we have constructed explicit classifiers (totally specified), and thus we do not need a training group. The whole group of one million of simulations was used for the inference issue.

Table 3 shows the inferential conclusions in relation to the problem of same performance of the classifiers. It contains representative  $p$ -values multiplied by 20 (to be compared with 0.05 because of Bonferroni correction). For each row ( $C_i$ ) and each column ( $C_j$ ), the table displays  $p$ -values  $\times 20$  for the null hypothesis  $H_0 : \widetilde{M}_{(C_j,T)} = \widetilde{M}_{(C_i,T)}$ , that is,  $H_0 : (C_j, T) \sim_{\widetilde{M}} (C_i, T)$ .

Thus, we conclude that there are not two classifiers at the same level in the alternative accumulated improvement curve stochastic order. Any two classifiers show a different performance to predict target  $T$ .

About the more efficient classifier problem, that is, the tests with null hypothesis  $H_0 : \widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$ , equivalently,  $H_0 : (C_j, T) \preceq_{\widetilde{M}} (C_i, T)$ , Table 4 contains the results. For each row ( $C_i$ ) and each column ( $C_j$ ), it includes the representative  $p$ -value multiplied by 20 (Bonferroni correction). As a consequence, our simulation study leads to the conclusion  $(C_2, T) \preceq_{\widetilde{M}} (C_3, T) \preceq_{\widetilde{M}} (C_4, T) \preceq_{\widetilde{M}} (C_1, T)$ .

Thus, the simulation study ‘corroborates’ the appropriate behaviour of the new technique for the comparison of classifiers.

To conclude, observe that  $T$  is positively regression dependent on  $C_1$ . Note that

$$P(T \leq t/C_1 = c) = \begin{cases} 1 & \text{if } t \geq 1, \\ 1 - \frac{e^c}{1+e^c} & \text{if } t \in [0, 1), \\ 0 & \text{if } t < 0, \end{cases}$$

and so,  $P(T \leq t/C_1 = c)$  is decreasing in  $c$  for any  $t$ . Applying Proposition 5.8, we conclude that  $\widetilde{M}_{(C_1,T)}$  is decreasing. Therefore, it is expected that the sample version of such a curve is decreasing. Observe that such a sample mapping is decreasing (see Figure 3).

**Acknowledgements.** The authors want to thank the Reviewer and the Editor for their interesting comments and suggestions, which have contributed to a substantial improvement of the manuscript. The authors are indebted to Ministerio de Economía y Competitividad (Spain) since this research is financed by Grants MTM2015-63971-P and MTM2017-83506-C2-2-P. We would like to express our deepest gratitude to Liberbank for its assistance, cooperation and allowing us the use of its databases.

**Disclosure statement.** No potential conflict of interest was reported by the authors.

## References

- Belzunce, F., Martínez-Riquelme, C., and Mulero, J. (2016). *An Introduction to Stochastic Orders*. Elsevier/Academic Press, Amsterdam.
- Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney.
- Breiman, L. (1992). *Probability*. Corrected reprint of the 1968 original. Classics in Applied Mathematics, 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Breiman, L. (2001). Random forests. *Mach. Learn.* **45** 5-32.
- Buckinx, W. and Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *Eur. J. Operat. Res.* **164** 252-268.
- Cai, J. and Wei, W. (2012). On the invariant properties of notions of positive dependence and copulas under increasing transformations. *Insurance Math. Econom.* **50** 43-49.
- Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*. Wiley Finance Series, John Wiley & Sons, Chichester.
- Cuadras, C. and Augé, J. (1981). A continuous general multivariate distribution and its properties. *Comm. Statist. A-Theory Methods* **10** 339-353.
- Esary, J.D. and Proschan, F. (1972). Relationships among some concepts of bivariate dependence. *Ann. Math. Statist.* **43** 651-655.

- Figini, S. and Giudici, P. (2010). Bayesian churn models. *Adv. Appl. Stat. Sci.* **1** 285-310
- Günther, C.C., Tvette, I.F., Aas, K., Sandnes, G.I. and Borgan, O. (2014). Modelling and predicting customer churn from an insurance company. *Scand. Actuar. J.* **1** 58-71.
- Hand, D.J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach. Learn.* **77** 103-123.
- Hand, D.J. (2010). Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Statist. Medicine* **29** 1502-1510.
- Hand, D.J. (2012). Assessing the performance of classification methods. *Internat. Statist. Rev.* **80** 400-414.
- Hand, D.J. and Anagnostopoulos, C. (2012). A better Beta for the H measure of classification performance. *Pattern Recognit. Lett.* **40** 41-46.
- Hand, D.J. and Anagnostopoulos, C. (2013). When is the area under the receiver characteristic curve an appropriate measure of classifier performance? *Pattern Recognit.* **80** 400-414.
- Hand, D.J. and Zhou, F. (2009). Evaluating models for classifying customers in retail banking collections. *J. Oper. Res. Soc.* **61** 1540-1547.
- Hung S., Yen D.C. and Wang H. (2006). Applying data mining to telecom churn management. *Expert. Syst. Appl.* **31** 515-524.
- Hwang H., Jung, T. and Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert. Syst. Appl.* **26** 181-188.
- Lee, W. (1999). Probabilistic analysis of global performances of diagnostic tests: interpreting the Lorenz curve-based summary measures. *Statist. Medicine* **18** 455-471.
- Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137-1153.
- Lloyd, C.J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *J. Amer. Statist. Assoc.* **93** 1356-1364.

- López-Díaz M.C., López-Díaz, M. and Martínez-Fernández, S. (2017). A stochastic comparison of customer classifiers with an application to customer attrition in commercial banking. *Scand. Actuar. J.* **7** 606-627.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester.
- Nelsen, R.B. (2006). *An introduction to copulas*. Second edition. Springer Series in Statistics, Springer, New York.
- Qi, J., Zhang, L., Liu, Y., Li, L., Zhou, Y., Shen, Y., Liang, L. and Li, H. (2009). ADTreesLogit model for customer churn prediction. *Ann. Oper. Res.* **168** 247-265.
- Rao, C.R. (1973). *Linear statistical inference and its applications*. Second edition. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York-London-Sydney.
- Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer, New York.
- Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- Wei, C. and Chiu, I. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert. Syst. Appl.* **23** 103-112.
- Yousef, W.A. (2013). Assessing classifiers in terms of the partial area under the ROC curve. *Comput. Statist. Data Anal.* **64** 51-70.

## Appendix: some proofs and auxiliary results

This appendix contains the proofs of some results in the manuscript.

*Proof of Proposition 3.4.* Note that  $F_X(F_X^{-1}(1-p)) = 1-p$  if and only if  $1-p \in \text{Im}(F_X)$  (see Proposition 1 in Shorack and Wellner (1986)), which implies that  $p^X = p$ . The converse is trivial.  $\square$

*Proof of Proposition 3.5.* Note that  $P(F_X(X) > 1-p^X) = 1 - P(F_X(X) \leq 1-p^X) = 1 - (1-p^X) = p^X$  since  $1-p^X \in \text{Im}(F_X)$  (see, for instance, Proposition 2 in Shorack and Wellner (1986)).  $\square$

The proof of Proposition 3.6 is based on the following result.

**Lemma A.1.** *Let  $X$  be a random variable and  $p \in (0, 1]$ . Then,  $p \in \text{Im}(F_X)$  and there exists  $\varepsilon > 0$  such that  $(p - \varepsilon, p) \cap \text{Im}(F_X) = \emptyset$ , if and only if,  $P(F_X(X) = p) > 0$ .*

*Proof.* Assume that  $p \in \text{Im}(F_X)$  and there exists  $\varepsilon > 0$  such that  $(p - \varepsilon, p) \cap \text{Im}(F_X) = \emptyset$ . Take  $a_p = F_X^{-1}(p)$ . Note that  $F_X(a_p) = p$  (see, for instance, Proposition 1 of Shorack and Wellner (1986)). Observe that  $F_X(a_p^-) < F_X(a_p)$ , thus  $0 < P(X = a_p) \leq P(F_X(X) = p)$ .

Conversely, let us suppose that  $P(F_X(X) = p) > 0$ . Therefore,  $p \in \text{Im}(F_X)$  and  $0 < P(X = F_X^{-1}(p))$  (see, for instance, Proposition 3 of Shorack and Wellner (1986)). So,  $F_X(F_X^{-1}(p)) - F_X((F_X^{-1}(p))^-) > 0$ . Thus,  $F_X$  is left discontinuous at  $F_X^{-1}(p)$ . If we take  $\varepsilon > 0$  with  $\varepsilon < P(X = F_X^{-1}(p))$ , then  $(p - \varepsilon, p) \cap \text{Im}(F_X) = \emptyset$ .  $\square$

*Proof of Proposition 3.6.* We know that  $p^X \leq p$ . Let us suppose that  $p^X < p$ . By Proposition 3.4, we have that  $1 - p \notin \text{Im}(F_X)$ , and thus  $[1 - p, 1 - p^X) \cap \text{Im}(F_X) = \emptyset$ . This is a contradiction with Lemma A.1, which reads that  $P(F_X(X) = 1 - p^X) > 0$ .  $\square$

We state the following lemma to give a proof of Proposition 4.5.

**Lemma A.2.** *Let  $X = (C_X, T_X)$  and  $X_n = (C_{X_n}, T_{X_n})$ ,  $n \in \mathbb{N}$ , be modelling vectors such that  $C_X$  is continuous and  $\lim_n C_{X_n} = C_X$  in distribution. For any  $p \in (0, 1]$ , we have that  $\lim_n p^{C_{X_n}} = p^{C_X}$ .*

*Proof.* The case  $p = 1$  is clear, note that  $1^C = 1$  for any random variable  $C$ . Let  $p \in (0, 1)$ . The distribution function  $F_{C_X}$  is continuous, therefore  $1 - p \in \text{Im}(F_{C_X})$ , and so  $p^{C_X} = p$ .

Let us suppose that the result is false. Then, there exists  $\varepsilon > 0$ , such that for any  $n_0 \in \mathbb{N}$ , there is  $n'_0 > n_0$  with  $|p^{C_{X_{n'_0}}} - p^{C_X}| > \varepsilon$ . It is not a restriction to suppose  $1 - p + \varepsilon < 1$  since  $1 - p < 1$ .

Let  $a \in \mathbb{R}$  such that  $F_{C_X}(a) = 1 - p + \frac{\varepsilon}{2}$ . The existence of  $a$  is because  $F_{C_X}$  is continuous and  $1 - p + \frac{\varepsilon}{2} < 1$ . Since  $F_{C_X}$  is continuous, by Polya's Theorem (see, for instance, Rao (1973)), we obtain that

$$\lim_n \|F_{C_{X_n}} - F_{C_X}\|_\infty = \lim_n \sup_{x \in \mathbb{R}} |F_{C_{X_n}}(x) - F_{C_X}(x)| = 0.$$

Thus, there exists  $m_0 \in \mathbb{N}$  with  $\|F_{C_{X_n}} - F_{C_X}\|_\infty < \varepsilon/4$  for any  $n \geq m_0$ . This implies that for any  $n \geq m_0$ , it holds that

$$1 - p + \frac{\varepsilon}{4} < F_{C_{X_n}}(a) < 1 - p + \frac{3}{4}\varepsilon.$$

As a consequence,  $F_{C_{X_n}}(a) \geq 1 - p^{C_{X_n}}$  since  $F_{C_{X_n}}(a) > 1 - p$  (recall that  $1 - p^{C_{X_n}}$  is the first value greater or equal to  $1 - p$  which belongs to  $\text{Im}(F_{C_{X_n}})$ ). Observe that taking  $n_0 \geq m_0$ , we obtain that

$$\varepsilon < |p^{C_{X_{n'_0}}} - p^{C_X}| = (1 - p^{C_{X_{n'_0}}}) - (1 - p) \leq F_{C_{X_{n'_0}}}(a) - (1 - p) < \frac{3}{4}\varepsilon,$$

which is a contradiction. Thus, that  $\lim_n p^{C_{X_n}} = p^{C_X}$ .  $\square$

*Proof of Proposition 4.5.* By hypothesis,  $\widetilde{M}_{X_m}(p) \leq \widetilde{M}_{Y_m}(p)$  for any  $p \in (0, 1]$  and any  $m \in \mathbb{N}$ .

Note that

$$\begin{aligned} \widetilde{M}_{X_m}(p) &= \frac{1}{p} (p^{C_{X_m}} P(T_{X_m} = 1/F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}) \\ &\quad + (p - p^{C_{X_m}}) P(T_{X_m} = 1/F_{C_{X_m}}(C_{X_m}) = 1 - p^{C_{X_m}})), \end{aligned}$$

and  $\widetilde{M}_X(p) = P(T_X = 1/F_{C_X}(C_X) > 1 - p)$  since  $p^{C_X} = p$ , observe that  $F_{C_X}$  is continuous.

Let us see that  $\lim_m \widetilde{M}_{X_m}(p) = \widetilde{M}_X(p)$  and  $\lim_m \widetilde{M}_{Y_m}(p) = \widetilde{M}_Y(p)$  for any  $p \in (0, 1]$ , which proves the result.

The case  $p = 1$  is trivial. Let  $p \in (0, 1)$ .

By Lemma A.2,  $\lim_m p^{C_{X_m}} = p$ . Thus, we should prove that

$$\lim_m P(T_{X_m} = 1/F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}) = P(T_X = 1/F_{C_X}(C_X) > 1 - p).$$

Observe that for  $m$  large enough,  $p^{C_{X_m}} < 1$ . Note that

$$\begin{aligned} &P(T_{X_m} = 1/F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}) \\ &= \frac{P(F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}, T_{X_m} = 1)}{P(F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}})}. \end{aligned}$$

Let us see that  $\lim_m P(F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}) = P(F_{C_X}(C_X) > 1 - p)$ .

By Theorem 5.5 in Billingsley (1968), we have that  $\lim_m F_{C_{X_m}}(C_{X_m}) = F_{C_X}(C_X)$  in distribution, note that  $F_{C_X}$  is continuous, and so, the sequence of distributions  $\{F_{C_{X_m}}\}_m$  tends to  $F_{C_X}$  uniformly. We have that

$$\begin{aligned} &P(F_{C_{X_m}}(C_{X_m}) \leq 1 - p^{C_{X_m}}) = F_{F_{C_{X_m}}(C_{X_m})}(1 - p^{C_{X_m}}) \\ &= F_{F_{C_{X_m}}(C_{X_m})}(1 - p^{C_{X_m}}) - F_{F_{C_X}(C_X)}(1 - p^{C_{X_m}}) + F_{F_{C_X}(C_X)}(1 - p^{C_{X_m}}). \end{aligned}$$

Thus,

$$|F_{F_{C_{X_m}}(C_{X_m})}(1 - p^{C_{X_m}}) - F_{F_{C_X}(C_X)}(1 - p^{C_{X_m}})| \leq \|F_{F_{C_{X_m}}(C_{X_m})} - F_{F_{C_X}(C_X)}\|_\infty,$$

which vanishes as  $m$  tends to infinity since  $\lim_m F_{C_{X_m}}(C_{X_m}) = F_{C_X}(C_X)$  in distribution and  $F_{F_{C_X}(C_X)}$  is continuous. On the other hand,

$$\lim_m F_{F_{C_X}(C_X)}(1 - p^{C_{X_m}}) = F_{F_{C_X}(C_X)}(1 - p)$$

since  $F_{F_{C_X}(C_X)}$  is continuous.

Now, let us prove that

$$\lim_m P(F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}, T_{X_m} = 1) = P(F_{C_X}(C_X) > 1 - p^{C_X}, T_X = 1)$$

for any  $p \in (0, 1)$ . Observe that

$$\lim_m F_{(F_{C_X}(C_X), T_X)}(1 - p^{C_{X_m}}, 1/2) = F_{(F_{C_X}(C_X), T_X)}(1 - p^{C_X}, 1/2)$$

since  $F_{(F_{C_X}(C_X), T_X)}$  is continuous in its first argument and  $\lim_m p^{C_{X_m}} = p^{C_X}$ .

Take the maps  $L_m, L : \mathbb{R} \rightarrow \mathbb{R}$ ,  $m \in \mathbb{N}$ , with  $L_m(x) = F_{(F_{C_{X_m}}(C_{X_m}), T_{X_m})}(x, 1/2)$  and  $L(x) = F_{(F_{C_X}(C_X), T_X)}(x, 1/2)$  for any  $x \in \mathbb{R}$ . They are increasing,  $L$  is continuous, and for any  $x \in \mathbb{R}$  we have that  $\lim_m L_m(x) = L(x)$  since  $(x, 1/2)$  is a continuity point of  $F_{(F_{C_X}(C_X), T_X)}$ . As a consequence, the sequence  $\{L_m\}_m$  converges uniformly to  $L$  in any compact interval. This implies that

$$\lim_m F_{(F_{C_{X_m}}(C_{X_m}), T_{X_m})}(1 - p^{C_{X_m}}, 1/2) - F_{(F_{C_X}(C_X), T_X)}(1 - p^{C_{X_m}}, 1/2) = 0$$

since  $\lim_n p^{C_{X_m}} = p$ . Thus, we have obtained that

$$\lim_m F_{(F_{C_{X_m}}(C_{X_m}), T_{X_m})}(1 - p^{C_{X_m}}, 1/2) = F_{(F_{C_X}(C_X), T_X)}(1 - p, 1/2).$$

Now, observe that

$$\begin{aligned} P(F_{C_{X_m}}(C_{X_m}) > 1 - p^{C_{X_m}}, T_{X_m} = 1) &= 1 - (P(F_{C_{X_m}}(C_{X_m}) \leq 1 - p^{C_{X_m}} \\ &\quad + P(T_{X_m} = 0) - F_{(F_{C_{X_m}}(C_{X_m}), T_{X_m})}(1 - p^{C_{X_m}}, 1/2)), \end{aligned}$$

and the same formula is satisfied by  $(F_{C_X}(C_X), T_X)$ , which leads to the proof of the result.  $\square$

We state the following lemma to prove Proposition 5.8.

**Lemma A.3.** *Let  $X = (C_X, T_X)$  be a modelling vector such that  $T_X$  is positively regression dependent on  $C_X$ . For any  $c_1, c_2 \in \mathbb{R}$ , with  $c_2 < c_1$ ,  $P(C_X > c_1) > 0$  and  $P(C_X \in (c_2, c_1)) > 0$ , it holds that*

- i)  $P(T_X = 1/C_X > c_1) \geq P(T_X = 1/C_X \geq c_1) \geq P(T_X = 1/C_X = c_1)$ ,
- ii)  $P(T_X = 1/C_X = c_1) \geq P(T_X = 1/c_1 > C_X > c_2)$ ,
- iii)  $P(T_X = 1/C_X \geq c_1) \geq P(T_X = 1/C_X > c_2)$ .

*Proof of Lemma A.3.* Since  $T_X$  is positively regression dependent on  $C_X$ ,  $P(T_X \leq t/C_X = c)$  is decreasing in  $c$  for any  $t$ . Thus,  $P(T_X = 1/C_X = c) = 1 - P(T \leq 0.5/C_X = c)$  is increasing in  $c$ . Moreover, for any Borel subset  $B \subset \mathbb{R}$ ,

$$P(C_X^{-1}(B), T_X = 1) = \int_B P(T_X = 1/C_X = s) dP_{C_X}.$$

Let us see i). It holds that

$$\begin{aligned} P(T_X = 1/C_X > c_1) &= \frac{P(C_X > c_1, T_X = 1)}{P(C_X > c_1)} \\ &= \frac{\int_{(c_1, +\infty)} P(T_X = 1/C_X = s) dP_{C_X}}{P(C_X > c_1)} \geq \frac{\int_{(c_1, +\infty)} P(T_X = 1/C_X = c_1) dP_{C_X}}{P(C_X > c_1)} \\ &= P(T_X = 1/C_X = c_1). \end{aligned}$$

On the other hand,  $P(T_X = 1/C_X \geq c_1)$

$$= \frac{P(C_X > c_1)}{P(C_X \geq c_1)} P(T_X = 1/C_X > c_1) + \frac{P(C_X = c_1)}{P(C_X \geq c_1)} P(T_X = 1/C_X = c_1).$$

Consider the mapping  $g : [0, 1] \rightarrow \mathbb{R}$  with  $g(\alpha) = \alpha P(T_X = 1/C_X > c_1) + (1 - \alpha) P(T_X = 1/C_X = c_1)$ . Note that  $g(\frac{P(C_X > c_1)}{P(C_X \geq c_1)}) = P(T_X = 1/C_X \geq c_1)$ . Since we have proved that  $P(T_X = 1/C_X > c_1) \geq P(T_X = 1/C_X = c_1)$ ,  $g$  is increasing. Therefore,  $g(0) \leq g(\frac{P(C_X > c_1)}{P(C_X \geq c_1)}) \leq g(1)$ , which proves i).

Let us analyze ii). Note that  $P(T_X = 1/c_1 > C_X > c_2)$

$$\begin{aligned} &= \frac{P(C_X \in (c_2, c_1), T_X = 1)}{P(C_X \in (c_2, c_1))} = \frac{\int_{(c_2, c_1)} P(T_X = 1/C_X = s) dP_{C_X}}{P(C_X \in (c_2, c_1))} \\ &\leq \frac{\int_{(c_2, c_1)} P(T_X = 1/C_X = c_1) dP_{C_X}}{P(C_X \in (c_2, c_1))} = P(T_X = 1/C_X = c_1). \end{aligned}$$

In relation to iii), observe that

$$P(T_X = 1/C_X > c_2) = \frac{P(C_X \geq c_1)}{P(C_X > c_2)} P(T_X = 1/C_X \geq c_1)$$

$$+\frac{P(C_X \in (c_2, c_1))}{P(C_X > c_2)}P(T_X = 1/C_X \in (c_2, c_1)).$$

Let  $g : [0, 1] \rightarrow \mathbb{R}$  with  $g(\alpha) = \alpha P(T_X = 1/C_X \geq c_1) + (1 - \alpha)P(T_X = 1/C_X \in (c_2, c_1))$ . By *i)* and *ii)*, this mapping is increasing. Thus, we obtain that  $g(\frac{P(C_X \geq c_1)}{P(C_X > c_2)}) \leq g(1)$ , which derives *iii)*.  $\square$

*Proof of Proposition 5.8.* Let  $p, p' \in (0, 1]$  with  $p \leq p'$ . Recall that  $p^{C_X} \leq p'^{C_X}$ .

Case 1) Suppose that  $p^{C_X} = p'^{C_X} = p_0$  for some  $p_0 \in (0, 1]$ . If  $p_0 = 1$ , the result is clear since  $p = p' = 1$ . Suppose that  $p_0 < 1$ . In this case,  $\widetilde{M}_X(p)$

$$= \frac{p_0}{p}P(T_X = 1/F_{C_X}(C_X) > 1 - p_0) + (1 - \frac{p_0}{p})P(T_X = 1/F_{C_X}(C_X) = 1 - p_0),$$

and  $\widetilde{M}_X(p')$

$$= \frac{p_0}{p'}P(T_X = 1/F_{C_X}(C_X) > 1 - p_0) + (1 - \frac{p_0}{p'})P(T_X = 1/F_{C_X}(C_X) = 1 - p_0).$$

By Lemma A.3 *i)*,  $P(T_X = 1/F_{C_X}(C_X) > 1 - p_0) \geq P(T_X = 1/F_{C_X}(C_X) = 1 - p_0)$ , and  $p_0/p \geq p_0/p'$ , which implies that  $\widetilde{M}_X(p) \geq \widetilde{M}_X(p')$ .

Case 2) Suppose that  $p^{C_X} < p'^{C_X}$ . Let  $m = P(F_{C_X}(C_X) = 1 - p^{C_X})$ . We have that

$$\begin{aligned} \widetilde{M}_X(p) &= \frac{p^{C_X}}{p}P(T_X = 1/F_{C_X}(C_X) > 1 - p^{C_X}) \\ &\quad + (1 - \frac{p^{C_X}}{p})P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X}) \\ &\geq \frac{p^{C_X}}{p^{C_X} + m}P(T_X = 1/F_{C_X}(C_X) > 1 - p^{C_X}) \\ &\quad + \frac{m}{p^{C_X} + m}P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X}) \\ &= P(T_X = 1/F_{C_X}(C_X) \geq 1 - p^{C_X}) \end{aligned}$$

since  $p \leq p^{C_X} + m$  and  $P(T_X = 1/F_{C_X}(C_X) > 1 - p^{C_X}) \geq P(T_X = 1/F_{C_X}(C_X) = 1 - p^{C_X})$ .

Since  $1 - p^{C_X} > 1 - p'^{C_X}$ , by Lemma A.3 *iii)*,  $P(T_X = 1/F_{C_X}(C_X) \geq 1 - p^{C_X}) \geq P(T_X = 1/F_{C_X}(C_X) > 1 - p'^{C_X})$ . Moreover, by Lemma A.3 *i)*,

$$\begin{aligned} P(T_X = 1/F_{C_X}(C_X) > 1 - p'^{C_X}) &\geq \frac{p'^{C_X}}{p'}P(T_X = 1/F_{C_X}(C_X) > 1 - p'^{C_X}) \\ &\quad + (\frac{p' - p'^{C_X}}{p'})P(T_X = 1/F_{C_X}(C_X) = 1 - p'^{C_X}) = \widetilde{M}_X(p'), \end{aligned}$$

which concludes the proof.  $\square$

*Proof of Proposition 5.9.* Let  $p \in (0, 1)$ . The continuity of  $C_X$  implies that

$$\begin{aligned} \widetilde{M}_X(p) &= P(T_X = 1 / F_{C_X}(C_X) > 1 - p) = \frac{P(F_{C_X}(C_X) > 1 - p, T_X = 1)}{P(F_{C_X}(C_X) > 1 - p)} \\ &= \frac{P(C_X > F_{C_X}^{-1}(1 - p), T_X = 1)}{P(C_X > F_{C_X}^{-1}(1 - p))} = \frac{\widehat{C}(\overline{F}_{C_X}(F_{C_X}^{-1}(1 - p)), q)}{\overline{F}_{C_X}(F_{C_X}^{-1}(1 - p))}, \end{aligned}$$

which leads to the result.  $\square$

*Proof of Proposition 5.10.* Let  $X, Y \in \Upsilon$ . Let  $\widehat{C}_{\theta_X}$  and  $\widehat{C}_{\theta_Y}$  be survival copulas of  $X$  and  $Y$ , respectively.

Let  $\underline{\theta} = \min\{\theta_X, \theta_Y\}$  and  $\overline{\theta} = \max\{\theta_X, \theta_Y\}$ . Let  $\underline{Z}$  and  $\overline{Z}$  be modelling vectors of  $\Upsilon$  such that their survival copulas are  $\widehat{C}_{\underline{\theta}}$  and  $\widehat{C}_{\overline{\theta}}$ , respectively.

It can be proved that  $\widehat{C}_{\theta_X}(p, q) \leq \widehat{C}_{\overline{\theta}}(p, q)$  and  $\widehat{C}_{\theta_Y}(p, q) \leq \widehat{C}_{\overline{\theta}}(p, q)$  for any  $p \in [0, 1]$ . Moreover,  $\widehat{C}_{\underline{\theta}}(p, q) \leq \widehat{C}_{\theta_X}(p, q)$  and  $\widehat{C}_{\underline{\theta}}(p, q) \leq \widehat{C}_{\theta_Y}(p, q)$  for any  $p \in [0, 1]$ . In accordance with Theorem 5.6, we deduce that  $\underline{Z} \preceq_{\widetilde{M}} X$ ,  $\underline{Z} \preceq_{\widetilde{M}} Y$ ,  $X \preceq_{\widetilde{M}} \overline{Z}$  and  $Y \preceq_{\widetilde{M}} \overline{Z}$ .

Let us see that  $\underline{Z}$  is the infimum of  $X$  and  $Y$  in the order  $\preceq_{\widetilde{M}}$ .

Let  $W \in \Upsilon$  such that  $W \preceq_{\widetilde{M}} X$  and  $W \preceq_{\widetilde{M}} Y$ . Let  $\widehat{C}_{\theta_W}$  be the survival copula of  $W$ . By Theorem 5.6, we have that  $\widehat{C}_{\theta_W}(p, q) \leq \widehat{C}_{\theta_X}(p, q)$  for any  $p \in [0, 1]$ . Then,  $q^{1-\theta_W} \leq q^{1-\theta_X}$  and so  $\theta_X \geq \theta_W$ . In the same way, we obtain that  $\theta_Y \geq \theta_W$ . Thus,  $\theta_W \leq \underline{\theta}$ . By Theorem 5.6,  $W \preceq_{\widetilde{M}} \underline{Z}$ . Therefore,  $\underline{Z}$  is the infimum of  $X$  and  $Y$  in the order  $\preceq_{\widetilde{M}}$ .

In identical manner, it is possible to see that  $\overline{Z}$  is the supremum of  $X$  and  $Y$  in the order  $\preceq_{\widetilde{M}}$ .  $\square$

*Proof of Proposition 5.11.* The result will be proved if we see that a survival copula of a continuity modelling vector of  $Z$ , say  $\widehat{C}_{CMV}^Z$ , is given by  $\widehat{C}_{CMV}^Z(u, v) = \min\{u, v\}$  for any  $u, v \in [0, 1]$ . Observe that any copula  $C$  satisfies that  $C(u, v) \leq \min\{u, v\}$  (Frechet-Hoeffding bounds for copulas). Thus, Theorem 5.5 would lead to the result.

It is not hard to see that

$$\widetilde{M}_Z(p) = \begin{cases} 1 & \text{if } 0 < p < q, \\ \frac{q}{p} & \text{if } q \leq p \leq 1. \end{cases}$$

Let  $Z' = (C'_Z, T'_Z)$  be a continuity modelling vector of  $Z$ . Observe that  $C'_Z \sim_{st}$

$U_{(0,1)}$  and  $T'_Z \sim_{st} \mathcal{B}(q)$ . Thus,

$$\overline{F}_{C'_Z}(k_1) = \begin{cases} 0 & \text{if } k_1 \geq 1, \\ 1 - x & \text{if } 0 \leq k_1 < 1, \\ 1 & \text{if } k_1 < 0, \end{cases} \quad \text{and} \quad \overline{F}_{T'_Z}(k_2) = \begin{cases} 0 & \text{if } k_2 \geq 1, \\ q & \text{if } 0 \leq k_2 < 1, \\ 1 & \text{if } k_2 < 0. \end{cases}$$

Take  $Z'$  the continuity modelling vector of  $Z$  constructed in Theorem 4.1. The survival function of  $Z'$  is

$$\overline{F}_{Z'}(k_1, k_2) = \begin{cases} 0 & \text{if } k_1 \geq 1 \text{ or } k_2 \geq 1, \\ (1 - k_1)M_Z(1 - k_1) & \text{if } 0 \leq k_1 < 1 \text{ and } 0 \leq k_2 < 1, \\ 1 - k_1 & \text{if } 0 \leq k_1 < 1 \text{ and } k_2 < 0, \\ q & \text{if } k_1 < 0 \text{ and } 0 \leq k_2 < 1, \\ 1 & \text{if } k_1 < 0 \text{ and } k_2 < 0, \end{cases}$$

for any  $(k_1, k_2) \in \mathbb{R}^2$ .

We should prove that  $\overline{F}_{Z'}(k_1, k_2) = \min\{\overline{F}_{C'_Z}(k_1), \overline{F}_{T'_Z}(k_2)\}$  for any  $k_1, k_2 \in \mathbb{R}$ .

Note that when  $0 \leq k_1 < 1$ , we have that

$$(1 - k_1)\widetilde{M}_Z(1 - k_1) = \begin{cases} 1 - k_1 & \text{if } 0 < 1 - k_1 < q, \\ q & \text{if } q \leq 1 - k_1 < 1, \end{cases}$$

which is equal to  $\min\{1 - k_1, q\} = \min\{\overline{F}_{C'_Z}(k_1), \overline{F}_{T'_Z}(k_2)\}$ .

Therefore,  $\widehat{C}_{CMV}^Z(u, v) = \min\{u, v\}$  is a survival copula of  $Z'$ , which proves the proposition.  $\square$

*Proof of Proposition 5.12.* It is sufficient to prove that a survival copula of a continuity modelling vector of  $W$ , is given by  $\widehat{C}_{CMV}^W(u, v) = \max\{u + v - 1, 0\}$  for any  $u, v \in [0, 1]$ . Note that any copula  $C$  satisfies that  $\max\{u + v - 1, 0\} \leq C(u, v)$  (Fréchet-Hoeffding bounds for copulas).

It can be seen that

$$\widetilde{M}_W(p) = \begin{cases} 0 & \text{if } 0 < p < 1 - q, \\ \frac{1}{p}(p + q - 1) & \text{if } 1 - q \leq p \leq 1. \end{cases}$$

Let  $W' = (C'_W, T'_W)$  be the continuity modelling vector of  $W$  defined in Theorem 4.1.

Observe that when  $0 \leq k_1 < 1$  and  $0 \leq k_2 < 1$ ,

$$\overline{F}_{W'}(k_1, k_2) = (1 - k_1)\widetilde{M}_W(1 - k_1) = \begin{cases} q - k_1 & \text{if } 0 < k_1 \leq q, \\ 0 & \text{if } q < k_1 < 1, \end{cases}$$

that is,  $\max\{q - k_1, 0\} = \max\{1 - k_1 + q - 1, 0\} = \max\{\overline{F}_{(1-T_Z)'}(k_1) + \overline{F}_{T_Z'}(k_2) - 1, 0\}$ , which leads to  $\overline{F}_{W'}(k_1, k_2) = \max\{\overline{F}_{C'_W}(k_1) + \overline{F}_{T'_W}(k_2) - 1, 0\}$  for any  $k_1, k_2 \in \mathbb{R}$ , and so to the result.  $\square$

*Proof of Proposition 5.13.* Let  $p_k = P(C > k)$ . Thus,  $1 - p_k \in \text{Im}(F_C)$  and  $p - p_k = P(C = k) = P(F_C(C) = 1 - p_k)$ . Observe that  $p_k = p^C$ . Note that the events  $(F_C(C) > 1 - p^C)$  and  $(C > k)$  are equal *a.s.*, and the same holds with the events  $(F_C(C) = 1 - p^C)$  and  $(C = k)$ . Therefore,

$$\begin{aligned} H_C(p) &= \frac{1}{q}(p^C P(T = 1/F_C(C) > 1 - p^C) + (p - p^C)P(T = 1/F_C(C) = 1 - p^C)) \\ &= \frac{1}{q}P(F_C(C) \geq 1 - p^C, T = 1) = \frac{1}{q}P(C \geq k, T = 1) = \text{Sens}(k). \end{aligned}$$

Thus,  $(p, H_C(p)) = (p, \text{Sens}(k)) = (P(C \geq k), \text{Sens}(k))$ .

In relation to *ii*), note that  $p^{C^C} = p^C$ , and so,  $H_C(p^C) = \frac{p^C}{q}P(T = 1/F_C(C) > 1 - p^C)$ . On the other hand,  $H_C(p^C + m) = \frac{1}{q}(p^C P(T = 1/F_C(C) > 1 - p^C) + mP(T = 1/F_C(C) = 1 - p^C))$ . Thus, we conclude that  $H_C(p) = (1 - \frac{p-p^C}{m})H_C(p^C) + \frac{p-p^C}{m}H_C(p^C + m)$ .  $\square$

## Tables

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$
$C_1$	-	0,00	0,28	0,00	0,00	0,00	0,00	0,00
$C_2$	-	-	0,00	0,00	0,00	0,00	0,00	0,00
$C_3$	-	-	-	0,00	0,00	0,00	0,00	0,00
$C_4$	-	-	-	-	0,00	0,00	0,00	0,00
$C_5$	-	-	-	-	-	0,03	0,00	0,00
$C_6$	-	-	-	-	-	-	0,00	0,00
$C_7$	-	-	-	-	-	-	-	0,17
$C_8$	-	-	-	-	-	-	-	-

Table 1: For each row ( $C_i$ ) and each column ( $C_j$ ), the representative  $p$ -value $\times 20$  (Bonferroni correction) for the null hypothesis  $\widetilde{M}_{(C_j,T)} = \widetilde{M}_{(C_i,T)}$  is displayed (Section 6).

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$
$C_1$	-	1,00	1,00	1,00	1,00	1,00	1,00	1,00
$C_2$	0,00	-	0,00	1,00	1,00	1,00	0,00	0,00
$C_3$	0,27	1,00	-	1,00	1,00	1,00	1,00	1,00
$C_4$	0,00	0,00	0,00	-	1,00	1,00	0,00	0,00
$C_5$	0,00	0,00	0,00	0,00	-	1,00	0,00	0,00
$C_6$	0,00	0,00	0,00	0,00	0,02	-	0,00	0,00
$C_7$	0,00	0,00	0,00	0,00	0,47	0,32	-	0,20
$C_8$	0,00	0,00	0,00	0,00	0,26	0,14	0,89	-

Table 2: For each row ( $C_i$ ) and each column ( $C_j$ ), the representative  $p$ -value $\times 20$  (Bonferroni correction) for the null hypothesis  $\widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$  is displayed (Section 6).

	$C_1$	$C_2$	$C_3$	$C_4$
$C_1$	-	0,00	0,00	0,00
$C_2$	-	-	0,00	0,00
$C_3$	-	-	-	0,00
$C_4$	-	-	-	-

Table 3: For each row ( $C_i$ ) and each column ( $C_j$ ), the representative  $p$ -value $\times 20$  (Bonferroni correction) for the null hypothesis  $\widetilde{M}_{(C_j,T)} = \widetilde{M}_{(C_i,T)}$  is displayed (Section 7).

	$C_1$	$C_2$	$C_3$	$C_4$
$C_1$	-	1,00	1,00	1,00
$C_2$	0,00	-	0,00	0,00
$C_3$	0,00	1,00	-	0,00
$C_4$	0,00	1,00	1,00	-

Table 4: For each row ( $C_i$ ) and each column ( $C_j$ ), the representative  $p$ -value $\times 20$  (Bonferroni correction) for the null hypothesis  $\widetilde{M}_{(C_j,T)} \leq \widetilde{M}_{(C_i,T)}$  is displayed (Section 7).

## Figures

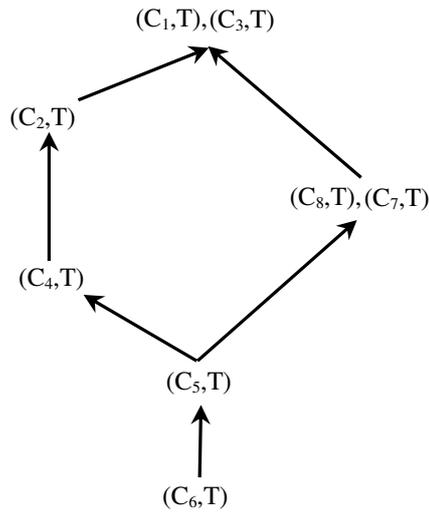


Figure 1: Diagram of classifiers for the alternative accumulated improvement curve stochastic order. An arrow from  $(C_j, T)$  to  $(C_i, T)$  means that  $C_i$  shows a better performance than  $C_j$  (Section 6).

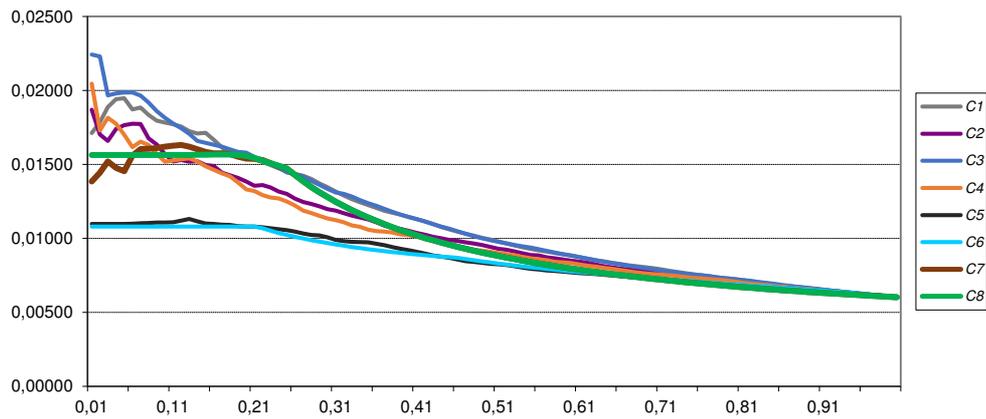


Figure 2: Sample alternative accumulated improvement curves of the classifiers, depicted with the whole validation group (Section 6).

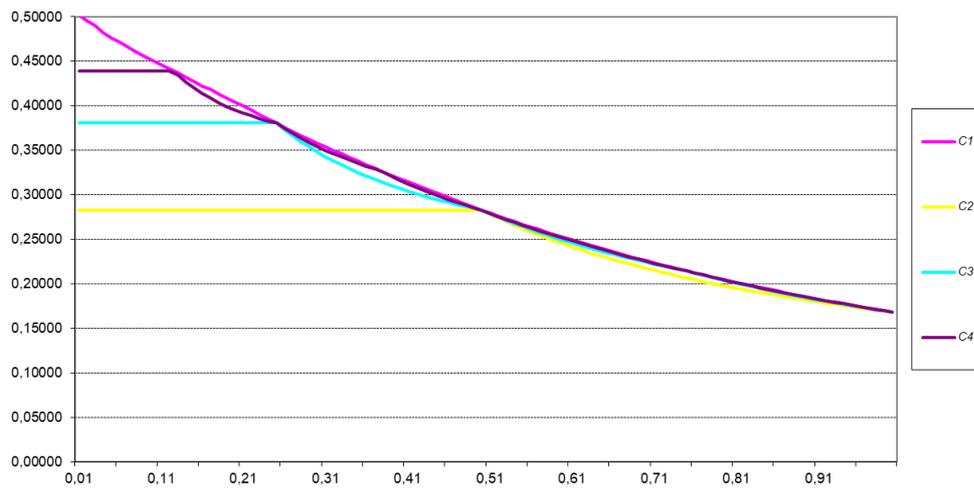


Figure 3: Sample alternative accumulated improvement curves of the classifiers, depicted with the whole set of simulations (Section 7).