A stochastic comparison of customer classifiers with an application to customer attrition in commercial banking

M.C. López-Díaz^a, M. López-Díaz^{b,*}, S. Martínez-Fernández^c

^{a}Departamento de Matemáticas, Universidad de Oviedo. C/Calvo Sotelo s/n. E-33007 Oviedo, Spain. e-mail:cld@uniovi.es
^{b}Departamento de Estadística e I.O. y D.M., Universidad de Oviedo. C/Calvo Sotelo s/n. E-33007 Oviedo, Spain. e-mail:mld@uniovi.es
^{c}Unidad de Modelos de Riesgos, Liberbank. Plaza de La Escandalera, 2. E-33003 Oviedo, Spain. e-mail:smartinezf@liberbank.es

Abstract

The classification of clients is an essential matter in commercial banking, insurance companies, electrical corporations, communication business, etc. Those companies frequently classify their customers by means of the information provided by the so-called classifier. Motivated by the need to compare systems of classification we introduce a new stochastic order which permits the comparison of classifiers. The stochastic order is analyzed in detail, providing characterizations and properties as well as connections with other stochastic orders and other classification systems. Such an order is applied to compare some classifiers used by a Spanish commercial banking to analyze the key problem of customer churn, obtaining conclusive results by means of real databases. Namely, the optimal classifier among them in the new stochastic order is obtained.

Keywords: churn rate, classifier, copula, logistic regression, stochastic order, target

---

*Corresponding author

1 Introduction

Customer classification plays a key role for commercial purposes in any marketing strategy. A suitable system of classification of customers is essential to develop efficient marketing programs.

A frequent situation is that in which customers are classified in two possible categories. That is very common in insurance companies or commercial banking. Data analysis units of banks often should classify their clients in a dichotomic system. For instance, will a customer purchase a credit card?, will a client abandon the bank?, will he acquire a mortgage?, will he be a defaulter?, etc. In fact, in many countries, commercial banks are obligated to classify their clients as "professional" or "retail client".

More examples can be found in insurance companies. These usually develop analysis of their clients, or potential clients, classifying them as future (or not future) purchasers of life insurances, health insurances, social insurances, third-party insurances, etc. Other examples are in relation to phone or electrical companies, which classify their former customers as returnable or not.

The classification of customers is very useful in developing marketing programs and provides information to launch campaigns or promotions among clients.

Classifications are frequently performed by means of a bidimensional random vector whose components are the so-called classifier and target. A target is a random variable which follows a Bernoulli distribution with parameter q, where q stands for the true proportion of customers in which the characteristic under study is satisfied. For instance, when a commercial bank analyzes which clients will purchase a credit card, q is the true and unknown proportion of customers which will buy the credit card. In this case the target takes value 1 at a client which will buy the credit card, and 0 if that client will not purchase it.

The value the target assigns to each client is commonly unknown (otherwise there would not exist uncertainty in the classification of clients), and is estimated using the information provided by the value the classifier takes at such a customer. For instance, if the classifier shows that the client has got into debt and his financial difficulties resulted in debt arrears, the bank probably will consider that he is not a potential purchaser of a pension and retirement plan, and in this case the bank will estimate that the value of the corresponding target at that person is 0.

Different techniques have been designed to construct classifiers, like those based on neuronal networks (see for instance Hwang *et al.* (2004) and Hung *et al.* (2006)), on decision trees (see Wei and Chiu (2002)), on random forest (see Breiman (2001) and Buckinx and Van den Poel (2005)), on logistic regression models (see Qi et al. (2009)), on Bayesian approaches (see Figini and Giudici (2010)), etc.

The comparison of such classifiers is an appealing topic. Detecting efficient classifiers to estimate a target is a key question to be successful with commercial campaigns among clients.

The main aim of this paper is to introduce a new mathematical framework to compare classifiers from a stochastic point view, analyze in detail such a model, and apply it to the comparison of classifiers used by a commercial banking to obtain conclusive results. Our mathematical model will be based on a stochastic order for classifiers and targets motivated by the problem described above. By means of such a stochastic order we develop a study where we compare multiple classifiers used in a Spanish bank to approach customer churn, obtaining that one of them (that based on logistic regression with all the variables and stepwise selection method) is the best in the new stochastic order to predict clients that are highly prone to abandon their bank. To the best of our knowledge stochastic orders have not been used to analyze the comparison of classifiers till this moment.

It is worth mentioning that customer churn has been widely analyzed in the last years. Customer churn prediction models try to find clients with high possibilities to break relationships with their companies (see Verbeke *et al.* (2011) for an overview on this topic). The churn prediction matter has been discussed in different fields, like in banking (see for instance Kumar and Ravi (2008) and Larivière and Van den Poel (2005)), insurance (see Morik and Kpcke (2004) and Günther *et al.* (2014)), telecommunications (see Hwang *et al.* (2004) and Hung *et al.* (2006)), retailing (Buckinx and Van den Poel (2005)), retail grocery sector (Miguéis *et al.* (2012)), etc.

The structure of the paper is as follows. In Section 2 we include the concepts and results needed for our analysis. The new stochastic order is introduced in Section 3, where a detailed analysis of the order is shown. In Section 4 we apply our mathematical model to some classifiers used in a Spanish commercial banking (Liberbank), using real data to compare eleven classifiers commonly applied by such a bank in the customer attrition problem. Conclusive results are derived by means of this study.

2   Preliminaries

Basic concepts and notations necessary for the development of the paper are included here.

Let W be a random variable or random vector. We will denote by F_W its distribution function, and by $\overline{F}_W$ its survival function.

When W is a random variable, the symbol F_W^{-1} will stand for the quantile function of W, that is, $F_W^{-1} : (0,1) \to \mathbb{R}$, with $F_W^{-1}(u) = \inf \{x \in \mathbb{R} : F_W(x) \geq u\}$ for any $u \in (0,1)$.

The usual Borel σ-algebra on $\mathbb{R}^d$ will be denoted by $\mathcal{B}_{\mathbb{R}^d}$.

If P is a probability on the measurable space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, and $h : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable

3

mapping, $P \circ h^{\star}$ will denote the probability given by $P \circ h^{\star}(B) = P(h^{\star}(B))$ for any $B \in \mathcal{B}_{\mathbb{R}^d}$, where $h^{\star}(B) = \{x \in \mathbb{R}^d : h(x) \in B\}$, that is, the preimage of B under h.

Let P be a probability on the measurable space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. A set $A \in \mathcal{B}_{\mathbb{R}^d}$ is said to be a continuity set for the probability P, if $P(\delta A) = 0$, where δA stands for the boundary of A.

The symbol $\sim$ will stand for the equality in distribution.

Two random vectors $(X_1, X_2, \ldots, X_d)$ and $(Y_1, Y_2, \ldots, Y_d)$ are said to belong to the same Frechet class if $X_i \sim Y_i$ for any $1 \leq i \leq d$, that is, marginal distributions of the same components are equal.

The dependence structure 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 $(U_1, U_2, \ldots, U_d)$ whose components follow uniform distribution on the interval $(0, 1)$.

If $X = (X_1, X_2, \ldots, X_d)$ is a random vector, then there exists C a copula such that $F_X = C(F_{X_1}, F_{X_2}, \ldots, F_{X_d})$. The reader is referred to Nelsen (2009) and Cherubini *et al.* (2004) for an introduction to the theory of copulas.

A key concept in this manuscript is the so-called stochastic order. Basically stochastic orders aim to rank probabilities in accordance with an appropriate criterion. Multiple criteria have been proposed in mathematical literature to order probabilities like variability, dispersion, concentration, residual life, etc.

Formally a stochastic order can be defined as a pre-order relation on a set of probabilities associated with a measurable space.

The reader is referred to the books Müller and Stoyan (2002), Shaked and Shanthikumar (2007) and Belzunce *et al.* (2015) for an introduction to the theory of stochastic orders from both theoretical an applied points of view.

Some stochastic orders which will appear throughout the paper are the following: let X and Y be random vectors, X is said to be smaller than Y in the

i) usual stochastic order, denoted by $X \preceq_{st} Y$, if $E(f(X)) \leq E(f(Y))$ for all increasing mappings $f : \mathbb{R}^d \to \mathbb{R}$ such that the above expectations exist, that is, for all mappings $f : \mathbb{R}^d \to \mathbb{R}$ with $f(x) \leq f(y)$ when $x \leq y$, where $\leq$ stands for the usual componentwise order of $\mathbb{R}^d$,

ii) upper orthant order, denoted by $X \preceq_{uo} Y$, if $\overline{F}_X(z) \leq \overline{F}_Y(z)$ for all $z \in \mathbb{R}^d$,

iii) lower orthant order, denoted by $X \preceq_{lo} Y$, if $F_X(z) \geq F_Y(z)$ for all $z \in \mathbb{R}^d$,

iv) positive quadrant dependent order, denoted by $X \preceq_{PQD} Y$, if $F_X(z) \leq F_Y(z)$ for all $z \in \mathbb{R}^2$, where $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ belong to the same Frechet class.

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

A random variable W is said to follow a Govindarajulu distribution with shape parameter $\beta > 0$, scale parameter $\sigma > 0$ and position parameter $\theta \geq 0$, if its quantile function is

given by $F_W^{-1}(u) = \theta + \sigma((\beta+1)u^\beta - \beta u^{\beta+1})$ for any $u \in (0,1)$ (see Govindarajulu (1977) for the definition of the family and Nair *et al.* (2012) for applications).

On the other hand, by "continuous random variable" we will mean a random variable whose distribution function is continuous.

3 The mathematical model: main results

In this section we introduce a new stochastic order to approach the problem of the comparison of classifiers described in the Introduction of the manuscript. Characterizations, properties and connections with other stochastic orders and other rating procedures are developed.

In the first place we introduce the concept of modelling vector, which will be the mathematical model the paper will be focused on.

Definition 3.1. *A bidimensional random vector $X = (C_X, T_X)$ is said to be a modelling vector if $T_X \sim \mathcal{B}(q)$. The random variables C_X and T_X are said to be the classifier and the target respectively.*

Roughly speaking, a classifier tries to "predict" the value of the target.

Definition 3.2. *Let $X = (C_X, T_X)$ be a modelling vector. The mapping $M_X : (0,1) \to \mathbb{R}$ 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.

The idea of this map is as follows. Given $p \in (0,1)$ we consider the smallest group containing at least the $100p\%$ of the customers with the largest values of the classifier C_X, and in such a subgroup we study the probability that the target takes value 1.

We should indicate that $P(C_X \geq F_{C_X}^{-1}(1-p)) > 0$ for any $p \in (0,1)$. As a consequence for any modelling vector $X = (C_X, T_X)$, the accumulated improvement curve M_X satisfies that

$$M_X(p) = \frac{P(C_X \geq F_{C_X}^{-1}(1-p), T_X = 1)}{P(C_X \geq F_{C_X}^{-1}(1-p))}$$

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

It is worth mentioning that in financial literature the accumulated improvement curve is sometimes defined as $P(T_X = 1/C_X \geq F_{C_X}^{-1}(1-p))/q$, that is, $M_X(p)/q$. As we will see

later that is irrelevant for our purposes since we will compare classifiers of targets with the same Bernoulli distribution.

A key applied problem in commercial banking is the comparison of classifiers. Based on this issue we introduce the following stochastic order to compare classifiers.

Definition 3.3. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be two modelling vectors with $T_X \sim T_Y \sim \mathcal{B}(q)$. It will be said that 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 $X \preceq_M Y$.*

The idea of the order is the following. Roughly speaking, the relation $X \preceq_M Y$ means that for any $p \in (0, 1)$, given the $100p\%$ of the customers 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 customers is greater (at least the same) with classifier C_Y than with classifier C_X.

For instance, let us consider the problem of the classification of customers of commercial banking as potential or not purchasers of a credit card. In this case the target, say T, is a random variable which takes 1 at a customer if he purchases the card and 0 in other case. If a bank make use of two modelling vectors $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ (in this case $T_X = T_Y = T$), the relation $X \preceq_M Y$ means that if the card is offered to the $100p\%$ of the clients with the largest values of the corresponding classifiers, the probability of having clients which acquire the card is greater (at least not smaller) when we consider the second classifier for any $p \in (0, 1)$.

From an applied point of view, in most of the cases T_X and T_Y are the same target. However there exist real-life problems in which targets could be different.

From now on we will assume that all the targets in the manuscript follow Bernoulli distribution with parameter q.

The new stochastic order provides a theoretical framework to compare efficiently classifiers.

3.1 Properties of the accumulated improvement curve

We start with the analysis of our mathematical model studying different results in relation to accumulated improvement curves.

Proposition 3.4. *Let $X = (C_X, T_X)$ be a modelling vector. It holds that $\lim_{p \to 1^-} M_X(p) = q$.*

Proof. Note that the events $C_X \geq F_{C_X}^{-1}(1 - p)$ are increasing in p. Therefore we have that

$$\lim_{p \to 1^-} P(C_X \geq F_{C_X}^{-1}(1 - p)) = \lim_{n \to \infty} P(C_X \geq F_{C_X}^{-1}(\frac{1}{n})) = P(\lim_{n \to \infty} (C_X \geq F_{C_X}^{-1}(\frac{1}{n})))$$

$$= P(\cup_{n=1}^{\infty}(C_X \geq F_{C_X}^{-1}(\frac{1}{n}))) = P(C_X \in supp\,(C_X)) = 1,$$

where *supp* stands for the support of the random variable. Reasoning in the same way, it is possible to obtain that

$$\lim_{p \to 1^-} P(C_X \geq F_{C_X}^{-1}(1-p), T_X = 1) = P(C_X \in supp\,(C_X), T_X = 1) = P(T_X = 1) = q.$$

Therefore we have that $\lim_{p \to 1^-} M_X(p) = q.$ $\qquad\square$

Observe that when p tends to 1, the event $(C_X \geq F_{C_X}^{-1}(1-p))$ tends to be satisfied by all the population, and so $M_X(p)$ tends to $P(T_X = 1)$.

Proposition 3.5. *Let* $X = (C_X, T_X)$ *be a modelling vector. If the classifier C_X is a continuous random variable, then the mapping M_X is continuous.*

Proof. Note that

$$M_X(p) = \frac{P(C_X \geq F_{C_X}^{-1}(1-p),\, T_X = 1)}{P(C_X \geq F_{C_X}^{-1}(1-p))}.$$

Let $p \in (0, 1)$. In accordance with formula (21) in Shorack and Wellner (1986), it holds that $C_X \geq F_{C_X}^{-1}(1-p)$ if and only if $F_{C_X}(C_X) \geq 1 - p$.

Moreover, if C_X is continuous, then $F_{C_X}(C_X) \sim U_{(0,1)}$. Thus $P(C_X \geq F_{C_X}^{-1}(1-p)) = p$, which is continuous at p.

In relation to the numerator of M_X, let us see that $P(T_X = 1, C_X \geq F_{C_X}^{-1}(1-p))$ is continuous at p, checking that it is right and left continuous.

Let $p_0 \in (0, 1)$. Let $\{p_n\}_n \subset (0, 1)$ be a decreasing sequence with $p_n > p_0$ for any $n \in \mathbb{N}$, and $\lim_n p_n = p_0$. Then

$$\lim_n P(C_X \geq F_{C_X}^{-1}(1-p_n), T_X = 1) = \lim_n P(F_{C_X}(C_X) \geq 1 - p_n, T_X = 1)$$

$$= P(\cap_{n=1}^{\infty}(F_{C_X}(C_X) \geq 1 - p_n, T_X = 1)) = P(F_{C_X}(C_X) \geq 1 - p_0, T_X = 1)$$

$$= P(C_X \geq F_{C_X}^{-1}(1-p_0), T_X = 1).$$

Therefore the numerator of M_X is right continuous.

Let $p_0 \in (0, 1)$. Let $\{p_n\}_n \subset (0, 1)$ be a increasing sequence with $p_n < p_0$ for any $n \in \mathbb{N}$, and $\lim_n p_n = p_0$. In this case

$$\lim_n P(C_X \geq F_{C_X}^{-1}(1-p_n), T_X = 1) = \lim_n P(F_{C_X}(C_X) \geq 1 - p_n, T_X = 1)$$

$$= P(\cup_{n=1}^{\infty}(F_{C_X}(C_X) \geq 1 - p_n, T_X = 1)) = P(F_{C_X}(C_X) > 1 - p_0, T_X = 1)$$

$$= P(F_{C_X}(C_X) \geq 1 - p_0, \, T_X = 1) = P(C_X \geq F_{C_X}^{-1}(1 - p_0), \, T_X = 1)$$

since the random variable C_X is continuous.

Therefore we obtain that M_X is continuous. $\qquad\square$

In general we can not guarantee that accumulated improvement curves are continuous for any modelling vector as the following example shows.

Example 3.6. Let us consider the modelling vector $X = (C_X, T_X)$ with $P(C_X = 0) = P(C_X = 1) = 1/2$. Let $a = P(T_X = 1/C_X = 0)$ and let $b = P(T_X = 1/C_X = 1)$, with $a, b \in [0, 1]$ and $a + b = 2q$.

In this case

$$F_{C_X}^{-1}(1 - p) = \begin{cases} 0 & \text{if } p \geq 1/2, \\ 1 & \text{if } p < 1/2. \end{cases}$$

It is not hard to prove that

$$M_X(p) = \begin{cases} b & \text{if } 0 < p < 1/2, \\ q & \text{if } 1/2 \leq p < 1. \end{cases}$$

Note that $\lim_{p \to 1^-} M_X(p) = q$ as Proposition 3.4 reads.

In general this mapping is neither continuous nor decreasing (increasing). For instance, if we take $b = 3q/4$ and $a = 5q/4$, where q is a value such that $5q/4 \leq 1$, then M_X is increasing, but not continuous. If we consider $a = 3q/4$ and $b = 5q/4$, where q is a value such that $5q/4 \leq 1$, then M_X is decreasing, but not continuous. If we take $a = b = q$, the mapping is continuous.

Next result reads that a classifier which is independent of the target does not provide us with information.

Proposition 3.7. *Let $X = (C_X, T_X)$ be a modelling vector. If C_X and T_X are independent, then $M_X(p) = q$ for any $p \in (0, 1)$.*

Proof. It is clear since $T_X \sim \mathcal{B}(q)$. $\qquad\square$

3.2 Basic properties of the stochastic order

In this subsection we analyze different properties of the accumulated improvement curve stochastic order.

Roughly speaking the first two results read that when classifiers are translated or multiplied by a positive scalar, accumulated improvement curves are not modified.

Proposition 3.8. *The stochastic order $\preceq_M$ is preserved under translations of the classifier.*

Proof. Let $X = (C_X, T_X)$ be a modelling vector and let $k \in \mathbb{R}$. We denote by $X \oplus k$ the modelling vector given by $X \oplus k = (C_X + k, T_X)$.

We have that $M_{X \oplus k}(p) = P(T_X = 1/C_X + k \geq F_{C_X+k}^{-1}(1-p))$. Now note that $C_X + k \geq F_{C_X+k}^{-1}(1-p)$ if and only if $F_{C_X+k}(C_X + k) \geq 1 - p$. But $F_{C_X+k}(x + k) = F_{C_X}(x)$, and so $M_{X \oplus k}(p) = P(T_X = 1/C_X \geq F_{C_X}^{-1}(1-p)) = M_X(p)$ for any $p \in (0,1)$. That is, $X \preceq_M X \oplus k$ and $X \oplus k \preceq_M X$. $\qquad \square$

Proposition 3.9. *The stochastic order $\preceq_M$ is preserved under scale changes of the classifier.*

Proof. Let $X = (C_X, T_X)$ be a modelling vector and $\lambda \in (0, +\infty)$. Let $\lambda \odot X = (\lambda C_X, T_X)$.

It holds that $P(\lambda C_X \geq F_{\lambda C_X}^{-1}(1-p)) = P(F_{\lambda C_X}(\lambda C_X) \geq 1-p) = P(F_{C_X}(C_X) \geq 1-p) = P(C_X \geq F_{C_X}^{-1}(1-p))$ for any $p \in (0,1)$.

In the same way $P(\lambda C_X \geq F_{\lambda C_X}^{-1}(1-p), T_X = 1) = P(C_X \geq F_{C_X}^{-1}(1-p), T_X = 1)$ for any $p \in (0,1)$.

Therefore $M_{\lambda \odot X} = M_X$, and so we have the result. $\qquad \square$

As a consequence of those results we obtain that the new stochastic order is not a partial order.

Proposition 3.10. *The relation $\preceq_M$ is a pre-order, but it does not satisfy the anti-symmetric property.*

Proof. It is clear that $\preceq_M$ is reflexive and transitive. However the anti-symmetric property is not satisfied. Observe that Proposition 3.8 says that $X \preceq_M X \oplus k$ and $X \oplus k \preceq_M X$ for any $k \in \mathbb{R}$, and obviously the modelling vectors $X \oplus k$ and X do not share the same distribution. $\qquad \square$

The new stochastic order is not preserved under weak convergence as the following example shows. However it is possible to state mild conditions which guarantee that preservation as we prove after the example.

Example 3.11. Let $X_m = (C_{X_m}, T_{X_m})$, $m \in \mathbb{N}$, be modelling vectors such that $P(C_{X_m} = 1, T_{X_m} = 1) = 3/8$, $P(C_{X_m} = 1, T_{X_m} = 0) = 1/8$, $P(C_{X_m} = 1 - 1/m, T_{X_m} = 1) = 1/8$, and $P(C_{X_m} = 1 - 1/m, T_{X_m} = 0) = 3/8$, for any $m \in \mathbb{N}$. Observe that $T_{X_m} \sim_{st} \mathcal{B}(1/2)$.

Note that for $0 < p < 1/2$, we have that $F_{C_{X_m}}^{-1}(1 - p) = 1$, while for $1/2 \leq p < 1$ it holds that $F_{C_{X_m}}^{-1}(1 - p) = 1 - 1/m$.

Thus it can be seen that

$$M_{X_m}(p) = \begin{cases} 3/4 & \text{if } 0 < p < 1/2, \\ 1/2 & \text{if } 1/2 \le p < 1, \end{cases}$$

for all $m \in \mathbb{N}$.

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 holds that $\lim_m X_m = X$ in the weak convergence.

Now consider $Y_m = (C_{Y_m}, T_{Y_m})$, $m \in \mathbb{N}$, modelling vectors such that $P(C_{Y_m} = 1, T_{Y_m} = 1) = 3/8$, $P(C_{Y_m} = 1, T_{Y_m} = 0) = 1/8$, $P(C_{Y_m} = 1/m, T_{Y_m} = 1) = 1/8$, and $P(C_{Y_m} = 1/m, T_{Y_m} = 0) = 3/8$, for any $m \in \mathbb{N}$.

In this case, we have that

$$M_{Y_m}(p) = \begin{cases} 3/4 & \text{if } 0 < p < 1/2, \\ 1/2 & \text{if } 1/2 \le p < 1, \end{cases}$$

for all $m \in \mathbb{N}$.

Let $Y = (C_Y, T_Y)$ be a modelling vector where $P(C_Y = 1, T_Y = 1) = 3/8$, $P(C_Y = 1, T_Y = 0) = 1/8$, $P(C_Y = 0, T_Y = 1) = 1/8$, and $P(C_Y = 0, T_Y = 0) = 3/8$. It holds that $\lim_m Y_m = Y$ in the weak convergence.

Now observe that $Y_m \preceq_M X_m$ for all $m \in \mathbb{N}$. However the relation $Y \preceq_M X$ is false since $M_X(p) = 1/2$ for any $p \in (0,1)$ (note that C_X and T_X are independent) while

$$M_Y(p) = \begin{cases} 3/4 & \text{if } 0 < p < 1/2, \\ 1/2 & \text{if } 1/2 \le p < 1. \end{cases}$$

Therefore the order $\preceq_M$ is not preserved under weak convergence.

In spite of that example, when classifiers are continuous the new order is preserved under the weak convergence as the following proposition asserts.

Proposition 3.12. *Let $X_m = (C_{X_m}, T_{X_m})$, $Y_m = (C_{Y_m}, T_{Y_m})$, $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$, with $m \in \mathbb{N}$, be modelling vectors. If $X_m \preceq_M Y_m$ for any $m \in \mathbb{N}$, $\lim_m X_m = X$ and $\lim_m Y_m = Y$ in the weak convergence, and C_X and C_Y are continuous random variables, then it holds that $X \preceq_M Y$.*

Proof. The condition $X_m \preceq_M Y_m$ implies that for any $p \in (0,1)$ and any $m \in \mathbb{N}$, $P(T_{X_m} = 1/C_{X_m} \ge F_{C_{X_m}}^{-1}(1-p)) \le P(T_{Y_m} = 1/C_{Y_m} \ge F_{C_{Y_m}}^{-1}(1-p))$, equivalently

$$\frac{P(C_{X_m} \ge F_{C_{X_m}}^{-1}(1-p), T_{X_m} = 1)}{P(C_{X_m} \ge F_{C_{X_m}}^{-1}(1-p))} \le \frac{P(C_{Y_m} \ge F_{C_{Y_m}}^{-1}(1-p), T_{Y_m} = 1)}{P(C_{Y_m} \ge F_{C_{Y_m}}^{-1}(1-p))}.$$

10

The hypothesis $\lim_m X_m = X$ and $\lim_m Y_m = Y$ in the weak convergence imply that $\lim_m C_{X_m} = C_X$ and $\lim_m C_{Y_m} = C_Y$ in the weak convergence since the weak convergence is preserved by the components of random vectors.

Thus it holds that $\lim_m F_{C_{X_m}}(x) = F_{C_X}(x)$ for any $x \in \mathbb{R}$ since F_{C_X} is continuous. Moreover the continuity of F_{C_X} implies that the sequence $\{F_{C_{X_m}}\}_m$ tends uniformly to F_{C_X} (see for instance Breiman (1968) or Pollard (1984)).

Therefore it is possible to apply Theorem 5.5 of Billingsley (1999) to obtain that that $\lim_m P_{C_{X_m}} \circ F^{\star}_{C_{X_m}} = P_{C_X} \circ F^{\star}_{C_X}$ and $\lim_m P_{C_{Y_m}} \circ F^{\star}_{C_{Y_m}} = P_{C_Y} \circ F^{\star}_{C_Y}$ in the weak convergence.

Observe that $P_{C_X} \circ F^{\star}_{C_X}$ is the probability induced by the random variable $F_{C_X}(C_X)$, and similarly with the rest of variables. Therefore we have that $\lim_m F_{C_{X_m}}(C_{X_m}) = F_{C_X}(C_X)$ and $\lim_m F_{C_{Y_m}}(C_{Y_m}) = F_{C_Y}(C_Y)$ in the weak convergence.

Thus we obtain that $\lim_m P(F_{C_{X_m}}(C_{X_m}) \geq 1 - p) = P(F_{C_X}(C_X) \geq 1 - p)$ for any $p \in (0,1)$ since the set $[1-p, \infty)$ is a continuity set for the probability induced by $F_{C_X}(C_X)$, note that the boundary of $[1-p, +\infty)$ is $\{1-p\}$ and $P(F_{C_X}(C_X) = 1-p) = 0$ since F_{C_X} is continuous.

Obviously we can derive the same conclusion with random variables $\{Y_m\}_m$ and Y.

Let us see now that for any $p \in (0,1)$ we have that

$$\lim_m P(F_{C_{X_m}}(C_{X_m}) \geq 1 - p, T_{X_m} = 1) = P(F_{C_X}(C_X) \geq 1 - p, T_X = 1). \qquad (3.1)$$

We know that $\lim_m X_m = X$ in the weak convergence and $\lim_m F_{C_{X_m}} = F_{C_X}$ in the uniform convergence.

As a consequence, if $g : \mathbb{R} \to \mathbb{R}$ is the identity function, we have that $\lim_m (F_{C_{X_m}}, g) = (F_{C_X}, g)$ uniformly in $\mathbb{R}^2$.

Applying Theorem 5.5 in Billingsley (1999) to the sequence $\{X_m\}_m$ which converges weakly to the vector X, we obtain that $\lim_m (F_{C_{X_m}}(C_{X_m}), T_{X_m}) = (F_{C_X}(C_X), T_X)$ in the weak convergence.

This implies formula (3.1) since $[1-p, \infty) \times \{1\}$ is a continuity set for the probability induced by $(F_{C_X}(C_X), T_X)$. Observe that F_{C_X} is continuous, and so $P(F_X(C_X) = 1 - p, T_X = 1) = 0$ for any $p \in (0,1)$.

Thus we have proved that for any $p \in (0,1)$ it holds that $\lim_m M_{X_m}(p) = M_X(p)$ and $\lim_m M_{Y_m}(p) = M_Y(p)$, and so derive the result. $\qquad \square$

3.3   Characterization results of the stochastic order

Now we provide some characterizations of the stochastic order $\preceq_M$.

The first result is based on quantile and distribution functions. The second characterization on the usual stochastic order.

Proposition 3.13. *Let* $X = (C_X, T_X)$ *and* $Y = (C_Y, T_Y)$ *be modelling vectors. Then* $X \preceq_M Y$ *if and only if*

$$F^{-1}_{T_X|F_{C_X}(C_X)\geq 1-p}\big(F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x)\big) \leq x \tag{3.2}$$

for any $p \in (0,1)$ *and any* x *such that* $F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x) \in (0,1)$.

Proof. In the first place, let us suppose that $X \preceq_M Y$. That is,

$$P(T_X = 1/C_X \geq F^{-1}_{C_X}(1-p)) \leq P(T_Y = 1/C_Y \geq F^{-1}_{C_Y}(1-p))$$

for any $p \in (0,1)$. This is the same as

$$P(T_X = 1/F_{C_X}(C_X) \geq 1 - p) \leq P(T_Y = 1/F_{C_Y}(C_Y) \geq 1 - p).$$

Since the marginal distributions of T_X and T_Y are Bernoulli, the above inequality can be written as

$$P(T_X > x/F_{C_X}(C_X) \geq 1 - p) \leq P(T_Y > x/F_{C_Y}(C_Y) \geq 1 - p)$$

for any $x \in \mathbb{R}$ and $p \in (0,1)$, that is,

$$P(T_X \leq x/F_{C_X}(C_X) \geq 1 - p) \geq P(T_Y \leq x/F_{C_Y}(C_Y) \geq 1 - p),$$

equivalently, $F_{T_X|F_{C_X}(C_X)\geq 1-p} \geq F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}$.

As a consequence we obtain that $F^{-1}_{T_X|F_{C_X}(C_X)\geq 1-p} \leq F^{-1}_{T_Y|F_{C_Y}(C_Y)\geq 1-p}$ and so we conclude that

$$F^{-1}_{T_X|F_{C_X}(C_X)\geq 1-p}\big(F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x)\big) \leq F^{-1}_{T_Y|F_{C_Y}(C_Y)\geq 1-p}\big(F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x)\big) \leq x$$

for any $p \in (0,1)$ and any x such that $F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x) \in (0,1)$.

Conversely, let us suppose that inequality (3.2) is satisfied. Since distribution functions are increasing we deduce that

$$F_{T_X|F_{C_X}(C_X)\geq 1-p}\big(F^{-1}_{T_X|F_{C_X}(C_X)\geq 1-p}\big(F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x)\big)\big) \leq F_{T_X|F_{C_X}(C_X)\geq 1-p}(x).$$

Now note that if F is a distribution function, then $F(F^{-1}(u)) \geq u$ for any $u \in (0,1)$, and thus we obtain that $F_{T_Y|F_{C_Y}(C_Y)\geq 1-p}(x) \leq F_{T_X|F_{C_X}(C_X)\geq 1-p}(x)$. This implies that

$$P(T_X = 1/F_{C_X}(C_X) \geq 1 - p) \leq P(T_Y = 1/F_{C_Y}(C_Y) \geq 1 - p)$$

for any $p \in (0,1)$, and so $X \preceq_M Y$. □

Proposition 3.14. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors. Then $X \preceq_M Y$ if and only if $T_X | (F_{C_X}(C_X) \geq 1 - p) \preceq_{st} T_Y | (F_{C_Y}(C_Y) \geq 1 - p)$ for any $p \in (0, 1)$.*

Proof. It is a consequence of Proposition 3.13 and Theorem 1.2.6 in Müller and Stoyan (2002) which reads that two random variables W_1 and W_2 satisfy $W_1 \preceq_{st} W_2$ if and only if $F_{W_1}^{-1}(F_{W_2}(x)) \leq x$. $\qquad\square$

The following result shows that when classifiers are continuous, the accumulated improvement curve stochastic order can be characterized by means of the upper orthant order considering appropriate transformations of modelling vectors.

Proposition 3.15. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors where the classifiers C_X and C_Y are continuous. Then it holds that $X \preceq_M Y$ if and only if $(F_{C_X}(C_X), T_X) \preceq_{uo} (F_{C_Y}(C_Y), T_Y)$.*

Proof. When C_X and C_Y are continuous, it holds that $F_{C_X}(C_X) \sim F_{C_Y}(C_Y) \sim U_{(0,1)}$. Therefore $P(C_X \geq F_{C_X}^{-1}(1-p)) = P(C_Y \geq F_{C_Y}^{-1}(1-p)) = p$ for any $p \in (0, 1)$.

As a consequence we obtain that $X \preceq_M Y$ if and only if

$$P(C_X \geq F_{C_X}^{-1}(1-p), T_X = 1) \leq P(C_Y \geq F_{C_Y}^{-1}(1-p), T_Y = 1)$$

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

This is the same as

$$P(F_{C_X}(C_X) \geq 1 - p, T_X = 1) \leq P(F_{C_Y}(C_Y) \geq 1 - p, T_Y = 1),$$

and because of the continuity of C_X and C_Y, this condition is equivalent to

$$P(F_{C_X}(C_X) > 1 - p, T_X = 1) \leq P(F_{C_Y}(C_Y) > 1 - p, T_Y = 1)$$

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

Since T_X and T_Y follow Bernoulli distribution, that is the same as

$$P(F_{C_X}(C_X) > 1 - p, T_X > x) \leq P(F_{C_Y}(C_Y) > 1 - p, T_Y > x)$$

for any $x \in \mathbb{R}$ and $p \in (0, 1)$. Note that if $x < 0$, both probabilities of the above expression are equal to p.

Observe that the above inequality of probabilities is also true when $p \leq 0$, both probabilities being equal to zero. If $p \geq 1$, the above probabilities are $P(T_X > x)$ and $P(T_Y > x)$ respectively, values which are equal for any $x \in \mathbb{R}$.

Therefore we have seen that $X \preceq_M Y$ if and only if $P(F_{C_X}(C_X) > y, T_X > x) \leq P(F_{C_Y}(C_Y) > y, T_Y > x)$ for any $x, y \in \mathbb{R}$, which is the same as $(F_{C_X}(C_X), T_X) \preceq_{uo} (F_{C_Y}(C_Y), T_Y)$. $\qquad\square$

Observe that when C_X and C_Y are continuous, then the random vectors $(F_{C_X}(C_X), T_X)$ and $(F_{C_Y}(C_Y), T_Y)$ belong to the same Frechet class. As a consequence of this we obtain the following corollary.

Corollary 3.16. *Let* $X = (C_X, T_X)$ *and* $Y = (C_Y, T_Y)$ *be modelling vectors where the classifiers* C_X *and* C_Y *are continuous. Then it holds that*

i) $X \preceq_M Y$ *if and only if* $(F_{C_X}(C_X), T_X) \preceq_{PQD} (F_{C_Y}(C_Y), T_Y)$,

ii) $X \preceq_M Y$ *if and only if* $(F_{C_Y}(C_Y), T_Y) \preceq_{lo} (F_{C_X}(C_X), T_X)$.

Proof. The results follow from p. 388 of Shaked and Shanthikumar (2007) and Proposition 3.15. $\qquad \square$

It is interesting to note that if C_X and C_Y are not continuous, the above results are not true as we show in the following example.

We should indicate that from an applied point of view, a great number of modelling vectors used in insurance firms or in commercial banking have classifiers which are not continuous, like for instance those based on decision trees.

Example 3.17. Let us consider the modelling vector $X = (C_X, T_X)$ with $P(C_X = 0) = P(C_X = 1) = 1/2$. Let $a = P(T_X = 1/C_X = 0)$ and let $b = P(T_X = 1/C_X = 1)$, with $a, b \in [0, 1]$ and $a + b = 2q$ (see Example 3.6).

It is not hard to prove that

$$P(T_X > x, F_{C_X}(C_X) > c) = \begin{cases} 1 & \text{if } x \in (-\infty, 0) \text{ and } c \in (-\infty, 1/2), \\ 1/2 & \text{if } x \in (-\infty, 0) \text{ and } c \in [1/2, 1), \\ 0 & \text{if } x \in (-\infty, 0) \text{ and } c \in [1, +\infty), \\ q & \text{if } x \in [0, 1) \text{ and } c \in (-\infty, 1/2), \\ b/2 & \text{if } x \in [0, 1) \text{ and } c \in [1/2, 1), \\ 0 & \text{if } x \in [0, 1) \text{ and } c \in [1, +\infty), \\ 0 & \text{if } x \in [1, +\infty) \text{ and } c \in \mathbb{R}. \end{cases}$$

The accumulated improvement curve of X is given by

$$M_X(p) = \begin{cases} b & \text{if } p \in (0, 1/2), \\ q & \text{if } p \in [1/2, 1). \end{cases}$$

Let $Y = (C_Y, T_Y)$ be a modelling vector with C_Y and T_Y independent random variables such that $C_Y = 1$ *a.s.*

In this case $P(T_Y > x, F_{C_Y}(C_Y) > c) = P(T_Y > x)P(F_{C_Y}(C_Y) > c)$. Note that $F_{C_Y}(C_Y) = 1$ a.s. Therefore

$$P(T_Y > x, F_{C_Y}(C_Y) > c) = \begin{cases} 0 & \text{if } x \in [1, +\infty) \text{ or } c \in [1, +\infty), \\ q & \text{if } x \in [0, 1) \text{ and } c \in (-\infty, 1), \\ 1 & \text{if } x \in (-\infty, 0) \text{ and } c \in (-\infty, 1). \end{cases}$$

It is clear that $M_Y(p) = q$ for any $p \in (0, 1)$ since C_Y and T_Y are independent (see Proposition 3.7).

Taking $q < b \leq 1$, we obtain that $Y \preceq_M X$.

On the other hand if $b < 2q$, then $P(T_X > x, F_{C_X}(C_X) > c) = b/2$ when $x \in [0, 1)$ and $c \in [1/2, 1)$, whereas $P(T_Y > x, F_{C_Y}(C_Y) > c) = q$. Therefore $(F_{C_Y}(C_Y), T_Y) \preceq_{uo} (F_{C_X}(C_X), T_X)$ is false.

Observe that there are values of a, b and q satisfying the above relations. It is sufficient to take $a = 0.3$, $b = 0.7$ and $q = 0.5$.

A consequence of the stochastic order $\preceq_M$ is now derived. The following lemma will be applied for that purpose.

Lemma 3.18. *Let $X = (C_X, T_X)$ be a modelling vector where C_X is a continuous random variable. Then the distribution of $T_X | C_X \geq x$ and $T_X | F_{C_X}(C_X) \geq F_{C_X}(x)$ are the same.*

Proposition 3.19. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors such that $X \preceq_M Y$, M_X is decreasing, C_X and C_Y are continuous and $C_Y \preceq_{st} C_X$. Then for any x such that $F_{C_X}(x), F_{C_Y}(x) \in (0, 1)$, it holds that $T_X | (C_X \geq x) \preceq_{st} T_Y | (C_Y \geq x)$.*

Proof. By Lemma 3.18 we have that

$$P(T_X > a/C_X \geq x) = P(T_X > a/F_{C_X}(C_X) \geq F_{C_X}(x)) = P(T_X > a/C_X \geq F_{C_X}^{-1}(F_{C_X}(x))).$$

Since M_X is decreasing and $F_{C_X} \leq F_{C_Y}$, we obtain that

$$P(T_X > a/C_X \geq F_{C_X}^{-1}(F_{C_X}(x))) \leq P(T_X > a/C_X \geq F_{C_X}^{-1}(F_{C_Y}(x))).$$

Using now the relation $X \preceq_M Y$,

$$P(T_X > a/C_X \geq F_{C_X}^{-1}(F_{C_Y}(x))) \leq P(T_Y > a/C_Y \geq F_{C_Y}^{-1}(F_{C_Y}(x)))$$

$$= P(T_Y > a/F_{C_Y}(C_Y) \geq F_{C_Y}(x)) = P(T_Y > a/C_Y \geq x).$$

As a consequence of this we obtain that

$$P(T_X \leq a/C_X \geq x) \geq P(T_Y \leq a/C_Y \geq x)$$

for any $a \in \mathbb{R}$, which concludes the proof. $\square$

3.4 Connections with copulas

Now we obtain conditions for the stochastic order $\preceq_M$ in relation to copulas of the modelling vectors.

Proposition 3.20. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors sharing a same copula C, with $C_X \preceq_{st} C_Y$. Let $f_{C,q} : (0,1) \to \mathbb{R}$ be the mapping given by $f_{C,q}(x) = (C(1-x, 1-q) - (1-q))/x$. Then,*

    i) if $f_{C,q}$ is increasing, it holds that $X \preceq_M Y$,

    ii) if $f_{C,q}$ is decreasing, it holds that $Y \preceq_M X$.

Proof. In accordance with the definition of the order $\preceq_M$ we should compare the values $M_X(p)$ and $M_Y(p)$ for any $p \in (0,1)$, that is,

$$\frac{P(C_X \geq F_{C_X}^{-1}(1-p), T_X = 1)}{P(C_X \geq F_{C_X}^{-1}(1-p))} \quad \text{and} \quad \frac{P(C_Y \geq F_{C_Y}^{-1}(1-p), T_Y = 1)}{P(C_Y \geq F_{C_Y}^{-1}(1-p))}$$

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

    By hypothesis, there exists C a common copula of $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$.

    Let $\widehat{C}$ be the survival copula of C, that is, the copula associated with the survival functions of the modelling vectors (C_X, T_X) and (C_Y, T_Y), namely, $\overline{F}_{(C_X, T_X)} = \widehat{C}(\overline{F}_{C_X}, \overline{F}_{T_X})$ and $\overline{F}_{(C_Y, T_Y)} = \widehat{C}(\overline{F}_{C_Y}, \overline{F}_{T_Y})$.

    If H is the copula of a bidimensional random vector (W_1, W_2) and $\widehat{H}$ is the survival copula of H, we have that

$$P(W_1 \geq w_1, W_2 > w_2) = P(\cap_{n=1}^{\infty}(W_1 > w_1 - \frac{1}{n}), W_2 > w_2) = \lim_n P(W_1 > w_1 - \frac{1}{n}, W_2 > w_2)$$

$$= \lim_n \widehat{H}(\overline{F}_{W_1}(w_1 - \frac{1}{n}), \overline{F}_{W_2}(w_2)) = \widehat{H}(\overline{F}_{W_1}(w_1^-), \overline{F}_{W_2}(w_2))$$

since copulas are continuous mappings (see for instance Nelsen (2009)) and survival functions are decreasing. In a similar way we obtain that $P(W_1 \geq w_1) = \widehat{H}(\overline{F}_{W_1}(w_1^-), 1)$.

    Thus in order to compare modelling vectors X and Y with the stochastic order $\preceq_M$, we should compare the values

$$\frac{\widehat{C}(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), \overline{F}_{T_X}(x))}{\widehat{C}(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), 1)} \quad \text{and} \quad \frac{\widehat{C}(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), \overline{F}_{T_Y}(x))}{\widehat{C}(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), 1)},$$

when for instance $x = 1/2$.

Note that if $x = 1/2$, then $\overline{F}_{T_X}(1/2) = P(T_X > 1/2) = P(T_X = 1) = q$, and the same formula is held with random variable T_Y.

As a consequence we should compare the values

$$\frac{\widehat{C}(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), q)}{\widehat{C}(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), 1)} \quad \text{and} \quad \frac{\widehat{C}(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), q)}{\widehat{C}(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), 1)}$$

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

Now if $C_X \preceq_{st} C_Y$, we have that $\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-) \leq \overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-)$ for any $p \in (0, 1)$.

Therefore if it is possible to state an inequality between the values $\widehat{C}(a, q)/\widehat{C}(a, 1)$ and $\widehat{C}(b, q)/\widehat{C}(b, 1)$ for any $a, b \in (0, 1)$ with $a \leq b$, we will obtain a relation in the stochastic order $\preceq_M$ between modelling vectors X and Y.

The copula associated with the survival function satisfies that $\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ for any $u, v \in [0, 1]$ (see Nelsen (2009)). Thus

$$\frac{\widehat{C}(a, q)}{\widehat{C}(a, 1)} = \frac{a + q - 1 + C(1 - a, 1 - q)}{a + 1 - 1 + C(1 - a, 0)} = \frac{q - 1}{a} + 1 + \frac{1}{a}C(1 - a, 1 - q) = f_{C,q}(a) + 1$$

since $C(1 - a, 0) = 0$.

If $f_{C,q}$ is increasing, then $f_{C,q}(a) + 1 \leq f_{C,q}(b) + 1$ for any $a, b \in (0, 1)$ with $a \leq b$, and so $M_X(p) \leq M_Y(p)$ for any $p \in (0, 1)$, that is, $X \preceq_M Y$.

On the other hand, if $f_{C,q}$ is decreasing, we obtain that $Y \preceq_M X$, which concludes the proof. $\qquad \square$

Next we compare different modelling vectors with the stochastic order $\preceq_M$ by applying the above result.

Example 3.21. Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors whose components are independent with $C_X \preceq_{st} C_Y$. In this case $C(u, v) = uv$ for any $u, v \in [0, 1]$. Thus the mapping $f_{C,q}$ is given by $f_{C,q}(x) = ((1 - x)(1 - q) - (1 - q))/x = q - 1$. This mapping is constant, therefore, increasing and decreasing.

As a consequence of Proposition 3.20 we obtain that $X \preceq_M Y$ and $Y \preceq_M X$. Note that this result was known since $M_X(p) = q = M_Y(p)$ for any $p \in (0, 1)$ (see Proposition 3.7). In fact for this result we do not need any relation between C_X and C_Y in the usual stochastic order.

Example 3.22. Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors sharing a same copula C, with $C_X \preceq_{st} C_Y$. Suppose that the common copula is $C : [0,1] \times [0,1] \to [0,1]$ with $C(u,v) = \min\{u,v\}$.

In this case it is possible to prove that $f_{C,q}(x) = \begin{cases} 0 & \text{if } x \leq q, \\ (q-x)/x & \text{if } x > q. \end{cases}$ Note that the derivative of the mapping $x \to (q-x)/x$ is negative, and so $f_{C,q}$ is decreasing. Thus we obtain that $Y \preceq_M X$.

Example 3.23. Let us consider now an example in which the mapping $f_{C,q}$ is increasing.

Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors sharing a same copula K, with $K(u,v) = u + v - 1 + C(1-u, 1-v)$, C being a copula, and $C_X \preceq_{st} C_Y$. Note that K is the survival copula of C.

In this case we conclude that $f_{K,q}(x) = -1 + C(x,q)/x$.

As a consequence, if the mapping $x \to C(x,q)/x$ is increasing, so is $f_{K,q}$. Let us consider the copula $C : [0,1] \times [0,1] \to [0,1]$ with $C(u,v) = uv/(1+(1-u)(1-v))$ (see Archimedean copulas in Nelsen (2009)).

It holds that $C(x,q)/x = q/(1+(1-q)(1-x))$, mapping which is increasing in x.

Then we obtain that $f_{K,q}$ is increasing. Applying Proposition 3.20 we conclude that $X \preceq_M Y$.

Example 3.24. We propose now an example in which the mapping $f_{C,q}$ is neither increasing nor decreasing.

Consider the Archimedean copula $C : [0,1] \times [0,1] \to [0,1]$ with $C(u,v) = \max\{1 - ((1-u)^\theta + (1-v)^\theta)^{1/\theta}, 0\}$, where $\theta \in (1, \infty)$.

In this case $f_{C,q}(x) = \left(\max\{1 - (x^\theta + q^\theta)^{1/\theta}, 0\} - (1-q)\right)/x$.

When the above maximum is 0, we obtain $f_{C,q}(x) = -(1-q)/x$, mapping which is increasing. Note that the above maximum is equal to 0 if and only if $x \geq (1-q^\theta)^{1/\theta}$.

If the involved maximum is greater than 0, that is, $x < (1-q^\theta)^{1/\theta}$, we have that $f_{C,q}(x) = (q - (x^\theta + q^\theta)^{1/\theta})/x$. For values of x with $x < (1-q^\theta)^{1/\theta}$, close enough to $(1-q^\theta)^{1/\theta}$, the derivative of $f_{C,q}$ is negative, and so that mapping is decreasing in such a part.

Thus, by Proposition 3.20 we are not able to derive a conclusion on the order $\preceq_M$ for modelling vectors $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ sharing copula C, with $C_X \preceq_{st} C_Y$.

Example 3.25. Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors with a common copula C given by $C(u,v) = uv/(u+v-uv)$, with $u, v \in (0,1)$.

Assume that the distributions of both classifiers belong to the so-called Govindarajulu family. Let $C_X \sim G(\beta, \sigma_1, \theta)$ and $C_Y \sim G(\beta, \sigma_2, \theta)$ with $\sigma_1 \leq \sigma_2$. It holds that $C_X \preceq_{st} C_Y$, note that $F_{C_X}^{-1}(u) \leq F_{C_Y}^{-1}(u)$ for any $u \in (0,1)$.

Consider Proposition 3.20 and the mapping $f_{C,q} : (0,1) \to \mathbb{R}$ defined there. It holds that $f_{C,q}(x) = -(1-q)^2/(2-q-x-(1-x)(1-q))$. It is not hard to prove that this map is decreasing. As a consequence of Proposition 3.20 we conclude that $Y \preceq_M X$.

The following results analyze the stochastic order $\preceq_M$ when we consider random vectors of the same Frechet class. Observe that no continuity conditions are imposed on the classifiers.

Proposition 3.26. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be modelling vectors in the same Frechet class whose copulas are C_1 and C_2 respectively. If*

  i) $\widehat{C}_1(a,q) \le \widehat{C}_2(a,q)$ *for any $a \in (0,1)$, then $X \preceq_M Y$,*

  ii) $\widehat{C}_1(a,q) \ge \widehat{C}_2(a,q)$ *for any $a \in (0,1)$, then $Y \preceq_M X$.*

Proof. Reasoning in the same way as in Proposition 3.20 we should compare the values $M_X(p)$ and $M_Y(p)$, that is,

$$\frac{\widehat{C}_1(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), q)}{\widehat{C}_1(\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-), 1)} \quad \text{and} \quad \frac{\widehat{C}_2(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), q)}{\widehat{C}_2(\overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-), 1)}$$

with $p \in (0,1)$.

Since $C_X \sim C_Y$, we obtain that $\overline{F}_{C_X}((F_{C_X}^{-1}(1-p))^-) = \overline{F}_{C_Y}((F_{C_Y}^{-1}(1-p))^-)$ for any $p \in (0,1)$ and so the comparison of $M_X(p)$ and $M_Y(p)$ reduces to the comparison of $\widehat{C}_1(a,q)/\widehat{C}_1(a,1)$ and $\widehat{C}_2(a,q)/\widehat{C}_2(a,1)$ with $a \in (0,1)$. Since $\widehat{C}_1(a,1) = \widehat{C}_2(a,1) = a$, we deduce that $M_X(p) \le M_Y(p)$ for any $p \in (0,1)$ if and only if $\widehat{C}_1(a,q) \le \widehat{C}_2(a,q)$ for any $a \in (0,1)$, which proves the results. $\square$

Next result provides maximal and minimal elements in the stochastic order $\preceq_M$ inside the same Frechet class.

Proposition 3.27. *Let $X = (C_X, T_X)$, $Y = (C_Y, T_Y)$ and $Z = (C_Z, T_Z)$ be modelling vectors in the same Frechet class whose copulas are C_1, C_2 and C_3 respectively. If copulas C_1 and C_3 satisfy that the associated survival copulas are $\widehat{C}_1(u,v) = \max\{u+v-1, 0\}$ and $\widehat{C}_3(u,v) = \min\{u,v\}$ respectively, then it holds that $X \preceq_M Y$ and $Y \preceq_M Z$.*

Proof. By Proposition 3.26 the result will be proved if we see that $\widehat{C}_1(a,q) \le \widehat{C}_2(a,q) \le \widehat{C}_3(a,q)$ for any $a \in (0,1)$.

Now note that $\widehat{C}_1$ and $\widehat{C}_3$ are the so-called universal Frechet-Hoeffding bounds for copulas, that is, for any copula $H : [0,1] \times [0,1] \to [0,1]$, it holds that $\widehat{C}_1(u,v) \le H(u,v) \le \widehat{C}_3(u,v)$ for any $u,v \in [0,1]$, which proves the result. $\square$

Now we derive the following consequences of this result.

Corollary 3.28. *Let $X = (C_X, T_X)$ be a modelling vector such that $C_X \sim_{st} \mathcal{B}(q)$. Then $X \preceq_M (T_X, T_X)$.*

Proof. Observe that for any $x, y \in \mathbb{R}$ we have that $P(T_X > x, T_X > y) = P(T_X > \max\{x, y\}) = \min\{P(T_X > x), P(T_X > y)\}$. Therefore the survival copula of the random vector (T_X, T_X) is $\widehat{C}(u, v) = \min\{u, v\}$, and so Proposition 3.27 leads to the result. $\square$

Corollary 3.29. *Let $X = (C_X, T_X)$ be a modelling vector such that $C_X \sim_{st} \mathcal{B}(q)$. Then $(1 - T_X, T_X) \preceq_M X$.*

Proof. In accordance with Proposition 3.27 it is sufficient to prove that the survival copula of $(1 - T_X, T_X)$ is $\widehat{C}(u, v) = \max\{u + v - 1, 0\}$.

That is, we should prove that $P(1 - T_X > x, T_X > y) = \widehat{C}(P(1 - T_X > x), P(T_X > y))$, where $x, y \in \mathbb{R}$.

Note that for any $x, y \in \mathbb{R}$, it holds that $P(1 - T_X > x, T_X > y) = P(y < T_X < 1 - x)$

$$
= \begin{cases}
0 & \text{if } y \geq 1 - x, \\
0 & \text{if } y < 1 - x \text{ and } 1 \leq x, \\
0 & \text{if } y < 1 - x, \ 0 \leq x < 1 \text{ and } y \geq 0, \\
1 - q & \text{if } y < 1 - x, \ 0 \leq x < 1 \text{ and } y < 0, \\
1 & \text{if } y < 1 - x, \ x < 0 \text{ and } y < 0, \\
q & \text{if } y < 1 - x, \ x < 0 \text{ and } y \in [0, 1), \\
0 & \text{if } y < 1 - x, \ x < 0 \text{ and } y \geq 1.
\end{cases}
$$

On the other hand, for any $x, y \in \mathbb{R}$ we have that $\widehat{C}(P(1 - T_X > x), P(T_X > y)) = \max\{P(1 - T_X > x) + P(T_X > y) - 1, 0\} = \max\{P(T_X < 1 - x) + P(T_X > y) - 1, 0\}$

$$
= \begin{cases}
0 & \text{if } 1 \leq x \text{ or } y \geq 1, \\
1 - q & \text{if } 0 \leq x < 1 \text{ and } y < 0, \\
0 & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y < 1, \\
1 & \text{if } x < 0 \text{ and } y < 0, \\
q & \text{if } x < 0 \text{ and } 0 \leq y < 1.
\end{cases}
$$

Therefore $P(1 - T_X > x, T_X > y) = \widehat{C}(P(1 - T_X > x), P(T_X > y))$ for any $x, y \in \mathbb{R}$ and so we have the result. $\square$

3.5 Comparison with other rating systems

Some rating systems used by companies are the so-called ROC indexes and CAP indexes, which are the areas under the ROC curves and under the CAP curves respectively (see for instance Bamber (1975) and Engelmann *et al.* (2003)). ROC and CAP are the acronyms of Receiver Operating Characteristic and Cumulative Accuracy Profile. We establish a relation between the system based on the accumulated improvement curve stochastic order and both indexes.

Consider a modelling vector (C, T) such that C assumes only the values 0 and 1, estimating T by the value of C. The table of possible outcomes is

	$T = 1$	$T = 0$
$C = 1$	TP	FP
$C = 0$	FN	TN

where TP means true positive, TN true negative, FP false positive and FN false negative. The sensitivity of the modelling vector is defined as the ratio between true positives and positive condition, that is, $P(C = 1, T = 1)/P(T = 1)$. The specificity as the ratio between true negatives and negative condition, that is, $P(C = 0, T = 0)/P(T = 0)$.

Let (C, T) be a general modelling vector. Given $k \in \mathbb{R}$, define the modelling vector (C_k, T) with $C_k = \begin{cases} 1 & \text{if } C \geq k, \\ 0 & \text{if } C < k. \end{cases}$

For each (C_k, T) consider its sensitivity and its specificity, we will denote them as $Sens(k)$ and $Spec(k)$ respectively, where $k \in \mathbb{R}$.

The ROC curve of the modelling vector (C, T) is given by $\{(1 - Spec(k), Sens(k)) \mid k \in \mathbb{R}\}$. The ROC index is defined as the area under the ROC curve.

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.

It is not hard to see the relation $(1 - q) \times$ ROC index$+(q/2) =$ CAP index.

The following result relates the new stochastic order with CAP curves.

Proposition 3.30. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be two modelling vectors where classifiers are continuous. If $X \preceq_M Y$ then the CAP curve of X is not greater than the CAP curve of Y.*

Proof. Let (C, T) be a modelling vector where C is continuous. Let $k \in \mathbb{R}$. Consider the modelling vector (C_k, T) defined above. Let $P(C \leq k) = 1 - p_k$. Suppose that $p_k \in (0, 1)$. Since C is continuous the events $(C \geq k)$ and $(F_C(C) \geq 1 - p_k)$ are equal *a.s.* Thus it holds that

$$Sens(k) = \frac{P(C \geq k, T = 1)}{P(T = 1)} = \frac{P(T = 1/F_C(C) \geq 1 - p_k)p_k}{q} = M_{(C,T)}(p_k)\frac{p_k}{q}.$$

Thus the CAP curve point associated with the value $k \in \mathbb{R}$ is $(p_k, M_{(C,T)}(p_k)\frac{p_k}{q})$.

As a consequence, since $M_X(p) \leq M_Y(p)$ for any $p \in (0,1)$ we obtain the result when $p_k \in (0,1)$. If $p_k = 0$ or $p_k = 1$, we have in the CAP curve the points $(0,0)$ and $(1,1)$ respectively. $\square$

Now we conclude that under mild conditions the accumulated improvement curve stochastic order implies that the corresponding ROC and CAP indexes are ordered.

Proposition 3.31. *Let $X = (C_X, T_X)$ and $Y = (C_Y, T_Y)$ be two modelling vectors where classifiers are continuous. If $X \preceq_M Y$ then*
   i) the CAP index of X is not greater than the CAP index of Y,
   ii) the ROC index of X is not greater than the ROC index of Y.

Proof. It is a consequence of the above proposition and the relation between the ROC and CAP indexes. $\square$

We should indicate that the converse of the above result is not true as the following example shows.

Consider a modelling vector $X = (C_X, T_X)$ satisfying that $P(C_X = 1) = P(C_X = 2) = P(C_X = 3) = 1/3$, $P(T_X = 1) = 0.4$, $P(T_X = 1|C_X = 3) = 0.6$, $P(T_X = 1|C_X = 2) = 0.3$ and $P(T_X = 1|C_X = 1) = 0.3$. On the other hand take the modelling vector $Y = (C_Y, T_Y)$ with $P(C_Y = 1) = P(C_Y = 2) = P(C_Y = 3) = 1/3$, $P(T_Y = 1) = 0.4$, $P(T_Y = 1|C_Y = 3) = 0.5$, $P(T_Y = 1|C_Y = 2) = 0.5$ and $P(T_Y = 1|C_Y = 1) = 0.2$.

It can be seen that both relations $X \preceq_M Y$ and $Y \preceq_M X$ are false, but ROC and CAP indexes of those classifiers are exactly the same.

4   An application to customer attrition in commercial banking

The mathematical model proposed and developed in Section 3 to compare classifiers has been applied to analyze the key problem of customer churn. This is a crucial matter for banks, insurance firms, telecommunication service providers, etc. Customer attrition rates are pivotal business indicators. The prediction of customer churn is a key issue to launch customer retention programs on those clients with high possibilities to abandon in order to retain them. It is interesting to observe that the cost of retaining a customer is much lower than the cost of acquiring a new client.

The mathematical tool introduced in this paper has been applied to compare different classifiers to detect clients that are highly prone to abandon their bank, in order to reduce churn rates.

Namely, our mathematical model has been used by a Spanish commercial banking, Liberbank (hereafter LBK), to compare eleven classifiers in relation to customer attrition. LBK is the result of the consolidation of three Spanish Savings Banks: Grupo Cajastur, Caja de Extremadura, and Caja Cantabria, to which they transfer the assets and liabilities of the banking business. LBK has a national presence and is a leader in retail markets of Asturias, Cantabria, Castilla-La Mancha and Extremadura, four regions of Spain.

The database used in this analysis is composed of clients of LBK that at 31 December 2013 they are aged between 18 and 80 years, have at least 500 euros of business balance (savings+debts), and an account balance greater than 0 euros. Employees of the bank, deceased, disabled and defaulter are not considered. In total 823.985 clients are included. The number of available variables of those clients is equal to 63. Among them, the variables number of financial products, age and business balance are considered the most important by LBK.

The target variable T is the abandonment of the bank. It was considered that a client had abandoned LBK when his business balance was 0 euros at 31 December 2014. The churn rate was 0.0386.

The set of customers containing 823.985 persons was divided into two groups, a training group with 523.985 clients, and a validation or testing group with 300.000 customers. The former was used for the development and estimation of the classifiers. The latter for our main aim, the comparison of classifiers to predict our target variable, that is, to predict if a client will abandon LBK.

The different models of classifiers considered in the analysis are the following:

- 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): Decision tree with all the variables, 4 branches per division, maximum depth of 8 and minimum size of leaf equals to 5.000 clients.

- Classifier 3 (C_3): Decision tree with all the variables, 2 branches per division, maximum depth of 3 and minimum size of leaf equals to 5.000 clients.

- Classifier 4 (C_4): Logistic regression with all the variables and stepwise selection method.

- Classifier 5 (C_5): Decision tree with the most important variables (number of financial products, age and business balance), 2 branches per division, maximum depth of 6 and minimum size of leaf equals to 5.000 clients.

- Classifier 6 (C_6): Logistic regression with the most important variables (number of financial products, age and business balance) and stepwise selection method.

- Classifier 7 (C_7): Decision tree with the variable number of financial products, 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 number of financial products, 2 branches per division, maximum depth of 2 and minimum size of leaf equals to 5.000 clients.

- Classifier 9 (C_9): Decision tree with the variable account balance, 2 branches per division, maximum depth of 2 and minimum size of leaf equals to 5.000 clients.

- Classifier 10 (C_{10}): Decision tree with the variable account balance, 2 branches per division, maximum depth of 3 and minimum size of leaf equals to 5.000 clients.

- Classifier 11 (C_{11}): Mesh method with the most important variables (number of financial products, age and business balance).

  In Classifier 11, the training group was divided by means of the deciles of the variables number of financial products, age and business balance, therefore generating 10^3 groups. Those groups whose sizes were less than 1000 were joined to groups with at least 1000 customers and with the nearest standard centroides.

Given any two classifiers, say C_i and C_j, we approach the following questions:

i) if C_i and C_j are equally efficient to predict abandonments (the value of T), that is, $(C_i, T) \preceq_M (C_j, T)$ and $(C_j, T) \preceq_M (C_i, T)$, equivalently, $M_{(C_i,T)} = M_{(C_j,T)}$,

ii) if C_i is more efficient than C_j to predict abandonments, namely, if the relation $(C_j, T) \preceq_M (C_i, T)$ is satisfied, that is to say, $M_{(C_j,T)} \leq M_{(C_i,T)}$.

For such a purpose and given any two classifiers, the validation group was divided at random 100 times into two groups of size 150.000. Each part of a division was assigned at random to one of the classifiers.

For each division and each classifier C, the mapping M was calculated in 20 points, that is, 20 values of p, from 0.05 to 1 with a step of 0.05. To that end, those customers in the corresponding part satisfying $F_C(C) \geq 1 - p$ were selected. The value of M at that p was obtained by means of the proportion of abandonments among those clients.

Thus, to compare any two classifiers C_i and C_j, we have 100 pairs of sample values of $M_{(C_i,T)}$ and $M_{(C_j,T)}$ at 20 points of p.

To approach question i), for each division of the validation group, and for each value of p we consider tests for the equality of the accumulated improvement curves at p of both classifiers. Namely, the null hypothesis of those tests is $H_0 : M_{(C_i,T)}(p) = M_{(C_j,T)}(p)$. Observe that this is a test for the comparison of two proportions.

Then for each division of the validation group we obtain 20 p-values associated with the 20 values of p (from 0.05 to 1 with a step of 0.05). As representative p-value of the comparison of $M_{(C_i,T)}$ and $M_{(C_j,T)}$ in that division, we take the smallest one among the 20 p-values That is, we consider the p-value showing more evidence that the relation $M_{(C_i,T)} = M_{(C_j,T)}$ is false.

In this way we have 100 p-values, one for each division of the validation group. As a summary p-value for the null hypothesis $M_{(C_i,T)} = M_{(C_j,T)}$ we consider the median p-value of that set.

Observe that to compare the accumulated improvement curves of two classifiers in one division of the validation group, we need to apply twenty tests (one for each possible value of p), thus we consider a Bonferroni correction to reduce the chances of obtaining false-positive results (type I errors).

We consider as a level of significance $\alpha = 0.05$. Therefore after Bonferroni correction we should compare the summary p-values with $0.05/20$.

Table 1 shows our conclusions in relation to question i). That table contains summary p-values multiplied by 20 (this is done for easy of reading, note that then they should be compared with the level of significance 0.05, roughly speaking, the Bonferroni correction is incorporated in the summary p-values).

For each row (C_i) and each column (C_j) we have the summary p-value$\times 20$ for the null hypothesis $M_{(C_j,T)} = M_{(C_i,T)}$.

As we can see, there is enough evidence to reject that any two classifiers are equal in the accumulated improvement curve stochastic order. Thus we conclude that there are not two classifiers equally efficient to predict customers that will abandon the bank.

The same procedure was followed to analyze question ii), that is, to test the hypothesis $M_{(C_j,T)} \leq M_{(C_i,T)}$. This means that classifier C_i is better that C_j to predict our target variable. Observe that in this case the null hypothesis of "intermediate tests" are $H_0 : M_{(C_j,T)}(p) \leq M_{(C_i,T)}(p)$.

Table 2 summarizes our conclusions. For each row (C_i) and each column (C_j) we have the summary p-value multiplied by 20 (Bonferroni correction) for the null hypothesis $M_{(C_j,T)} \leq M_{(C_i,T)}$.

As a consequence of the preceding inferential procedure we obtain the following relations among the classifiers for the accumulated improvement stochastic order (relations which can be deduced by transitivity are not included):

- $(C_3, T) \preceq_M (C_5, T) \preceq_M (C_1, T) \preceq_M (C_2, T) \preceq_M (C_4, T)$

- $(C_5, T) \preceq_M (C_6, T) \preceq_M (C_2, T)$

- $(C_3, T) \preceq_M (C_{11}, T) \preceq_M (C_6, T)$

- $(C_{10}, T) \preceq_M (C_9, T) \preceq_M (C_7, T) \preceq_M (C_{11}, T)$

- $(C_8, T) \preceq_M (C_7, T) \preceq_M (C_1, T)$

Observe that the above relations permit to consider a Hasse diagram (see for instance Schröder (2003), Harzheim (2005) and Neggers and Kim (1998)) to visualize clearly the relation among classifiers.

In the Hasse diagram the different classifiers are represented in the plane. If the relation $(C_j, T) \preceq_M (C_i, T)$ is satisfied, then the point of the modelling vector (C_i, T) has a larger y-coordinate that the point of the classifier (C_j, T), and we draw a line between the points of (C_j, T) and (C_i, T).

The corresponding Hasse diagram appears in Figure 1.

Thus we can conclude that among the eleven classifiers considered in the analysis, there is one maximal element in the accumulated improvement curve stochastic order, which is the classifier constructed by means of the logistic regression with all the variables and stepwise selection method. This has been a conclusive result for LBK in order to predict customers "predisposed" to abandon the bank and launch on them retention marketing campaigns.

We also include a graphical representation of the empirical accumulated improvement curves of the different classifiers (see Figure 2). That was obtained by means of the whole validation group. We can observe that the curve associated with classifier C_4 is above the rest of the curves. Those mappings were calculated at the above mentioned 20 values of p. Interpolation was considered to plot curves. "Steps" suggest discontinuity points.

To conclude this section, it is worth mentioning that the proposed technique in this manuscript is currently applied for LBK to compare classifiers for different targets like the purchase of a credit card, investment funds, acquisition of a bank loan, etc.

We should indicate that all the calculations were executed with SAS software. The data processing, the calculations of the accumulated improvement curves and the statistical inference were executed with SAS Base and SAS Stat. On the other hand, classifiers were constructed by means of SAS Enterpriser Miner.

References

Bamber, D. (1975) The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.*, 12, 387-415.

Belzunce, F., Riquelme, C., Mulero, J. (2015) *An Introduction to Stochastic Orders.* Academic Press.

Billingsley, P. (1999) *Convergence of Probability Measures.* New-York, Wiley Series in Probability and Statistics, John Wiley & Sons.

Breiman, L. (1968) *Probability.* Mass.-London-Don Mills, Ontario, Addison-Wesley Publishing Company, Reading.

Breiman, L. (2001) Random forests. *Mach. Learn.*, 45, 5-32.

Buckinx, W., 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.

Cherubini, U., Luciano, E., Vecchiato, W. (2004) *Copula methods in finance.* Chichester, Wiley Finance Series, John Wiley & Sons.

Engelmann, B., Hayden, E., Tasche, D. (2003) Measuring the Discriminative Power of Rating Systems. Discussion paper Series 2: Banking and Financial Supervision No 01/2003.

Figini, S., Giudici, P. (2010) Bayesian churn models. *Adv. Appl. Stat. Sci.*, 1, 285-310.

Govindarajulu, Z. (1977) A class of distributions useful in life testing and reliability with applications to nonparametric testing, in The theory and applications of reliability, with emphasis on Bayesian and nonparametric methods. New York, Academic Press, Vol. I, pp. 109-129. (Conf., Univ. South Florida, Tampa, Fla., (1975))

Günther, C.C., Tvete, Ingunn F., Aas, K., Sandnes, G.I.; Borgan, O. (2014) Modelling and predicting customer churn from an insurance company. *Scand. Actuar. J.*, 2014, 1, 58-71.

Harzheim, E. (2005) *Ordered sets.* Advances in Mathematics. New York, Springer.

Hung S., Yen D.C, Wang H. (2006) Applying data mining to telecom churn management. *Expert. Syst. Appl.*, 31, 515-524.

Hwang H., Jung, T., 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.

Kumar, D.A., Ravi, V. (2008) Predicting credit card customer churn in banks using data mining. *Int. J. Data Anal. Tech. Strat.*, 1, 4-28.

Larivière, B. Van den Poel, D. (2005) Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert. Syst. Appl.*, 29, 472484.

Miguéis, V.L., Van den Poel, D., Camanho, A.S., Falco e Cunha, J. (2012) Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences. *Adv. Data Anal. Classif.*, 6, 337-353.

Morik, K., Kpcke, H. (2004) Analysing customer churn in insurance data: a case study. In: Boulicaut, J.F., Esposito, F., Giannotti, F., Pedreschi, D.(eds) Knowledge discovery in databases. Lecture Notes in Computer Science, vol 3202. Springer, Italy, 325-336.

Müller, A., Stoyan, D. (2002) *Comparison Methods for Stochastic Models and Risks.* Chichester, John Wiley & Sons.

Nair, N.U., Sankaran, P.G., Vineshkumar, B. (2012) Characterization of distributions by properties of truncated Gini index and mean difference. *Metron*, 70, 173-191.

Neggers, J., Kim, H.S. (1998) *Basic posets.* Singapore, World Scientific Publishing.

Nelsen, R.B. (2009) *An introduction to copulas.* New York, Second edition. Springer Series in Statistics, Springer.

Pollard, D. (1984) *Convergence of stochastic processes.* New York, Springer Series in Statistics, Springer-Verlag.

Qi, J., Zhang, L., Liu, Y., Li, L., Zhou, Y., Shen, Yao., Liang, L., Li, H. (2009) ADTreesLogit model for customer churn prediction. *Ann. Oper. Res.*, 168, 247-265.

Schröder, B.S.W. (2003) *Ordered sets. An introduction.* Boston-Basel-Berlin, Brikhäuser.

Shaked, M., Shanthikumar, J.G. (2007) *Stochastic Orders.* New York, Springer.

Shorack, G.R., Wellner, J.A. (1986) *Empirical processes with applications to statistics.* New York, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Verbeke, W., Martens, D., Mues, C., Baesens, B. (2011) Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert. Syst. Appl.*, 38, 2354-2364.

Wei, C., Chiu, I. (2002) Turning telecommunications call details to churn prediction: a data mining approach. *Expert. Syst. Appl.*, 23, 103-112.

Tables

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
C1	.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C2	.	.	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C3	.	.	.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C4	.	.	.	.	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C5	.	.	.	.	.	0,00	0,00	0,00	0,00	0,00	0,00
C6	.	.	.	.	.	.	0,00	0,00	0,00	0,00	0,01
C7	.	.	.	.	.	.	.	0,00	0,00	0,00	0,00
C8	.	.	.	.	.	.	.	.	0,00	0,00	0,00
C9	.	.	.	.	.	.	.	.	.	0,02	0,00
C10	.	.	.	.	.	.	.	.	.	.	0,00
C11	.	.	.	.	.	.	.	.	.	.	.

Table 1: For each row (C_i) and each column (C_j) the summary p-value$\times 20$ for the null hypothesis $M_{(C_j,T)} = M_{(C_i,T)}$ is displayed.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
C1	.	0,00	1,00	0,00	1,00	0,00	0,34	0,21	1,00	1,00	0,00
C2	1,00	.	1,00	0,01	1,00	1,00	1,00	1,00	1,00	1,00	1,00
C3	0,00	0,00	.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C4	1,00	1,00	1,00	.	1,00	1,00	1,00	1,00	1,00	1,00	1,00
C5	0,00	0,00	0,25	0,00	.	0,00	0,00	0,04	0,00	0,28	0,00
C6	0,00	0,00	1,00	0,00	1,00	.	1,00	1,00	1,00	1,00	0,16
C7	0,00	0,00	0,00	0,00	0,00	0,00	.	1,00	0,07	0,91	0,00
C8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	.	0,00	0,00	0,00
C9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	.	1,00	0,00
C10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	.	0,00
C11	0,00	0,00	0,19	0,00	0,00	0,02	0,18	0,23	1,00	1,00	.

Table 2: For each row (C_i) and each column (C_j) the summary p-value$\times 20$ for the null hypothesis $M_{(C_j,T)} \leq M_{(C_i,T)}$ is displayed.
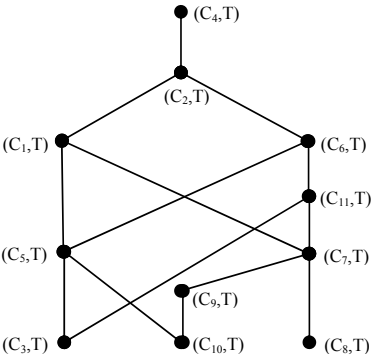
Figures

Figure 1: Hasse diagram of classifiers for the accumulated improvement stochastic order.
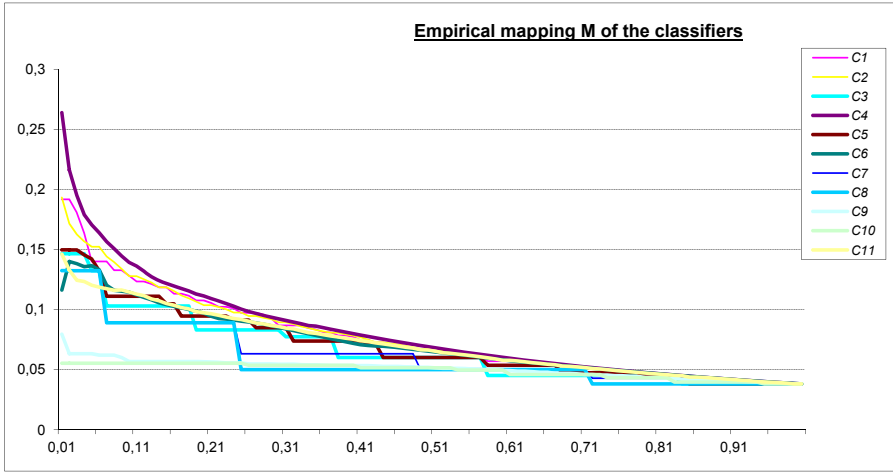

Figure 2: Empirical accumulated improvement curves of the classifiers obtained by means of the whole validation group.