Are all distributions distorted?

Marco Capaldo

Institute of Statistics, RWTH Aachen University, Germany

Meeting
“Stochastic Models in Biomathematics and Applications”
SMOD 2026

20-21 January 2026, Salerno, Italy

Table of contents

Table of contents

Table of contents

Table of contents

Table of contents

Table of contents

Reference paper

The talk is based on the following paper:

Capaldo M., Arevalillo J.M., Navarro J. (2025). Distorted distributions and ROC curves. Scandinavian Journal of Statistics 52, 1786–1815.

Table of Contents

Notation

Let X be a random variable.

- $F(t) = \Pr(X \leq t)$, $t \in \mathbb{R}$, denotes its cumulative distribution function (CDF);
- $\overline{F} = 1 - F$ denotes its survival function (SF);
- $F^{-1}(u) = \sup\{t\colon F(t) \leq u\}$, $u \in [0,1]$, is the quantile function;
- $\overline{F}^{-1}(u) = F^{-1}(1-u)$, for all $u \in [0,1]$;
- $\mu = \mathrm{E}(X) = -\int_{-\infty}^{0} F(t)\mathrm{d}t + \int_{0}^{+\infty} \overline{F}(t)\mathrm{d}t < +\infty$.

If F is absolutely continuous, then

- $f = F'$ is the probability density function (PDF);
- $\lambda = f/\overline{F}$ denotes the hazard rate (HR);
- $\tau = f/F$ denotes the reversed hazard rate (RHR);
- $\eta = -f'/f$ denotes the Glaser's function, provided that f is differentiable.

Definition

We say that X is smaller than Y in the

- usual stochastic order, denoted by $X \leq_{st} Y$, if $\overline{F}_X(t) \leq \overline{F}_Y(t)$ holds for all t. If there is equality in law, then we write $X =_{st} Y$;
- hazard rate order, denoted by $X \leq_{hr} Y$, if $\overline{F}_Y(t)/\overline{F}_X(t)$ is increasing in t;
- reversed hazard rate order, denoted by $X \leq_{rhr} Y$, if $F_Y(t)/F_X(t)$ is increasing in t;
- likelihood ratio order, denoted by $X \leq_{lr} Y$, if $f_Y(t)/f_X(t)$ is increasing in t in the union of their supports, provided that F_X and F_Y are absolutely continuous.

Aging classes

Definition

A random variable X with an absolutely continuous distribution is said to have the

- increasing hazard rate (IHR) property if its HR λ is increasing. If λ is decreasing, then X is said to have the decreasing hazard rate (DHR) property;
- increasing reversed hazard rate (IRHR) property if its RHR τ is increasing. If τ is decreasing, then X is said to have the decreasing reversed hazard rate (DRHR) property;
- increasing in likelihood ratio (ILR) property if its Glaser's function η is increasing. If η is decreasing, then X is said to have the decreasing in likelihood ratio (DLR) property;
- increasing density (ID) property if f is increasing. If f is decreasing, then X is said to have the decreasing density (DD) property.

Table: Relationships among the stochastic orders introduced above.

$X \leq_{lr} Y$	$\Rightarrow$	$X \leq_{hr} Y$
$\Downarrow$		$\Downarrow$
$X \leq_{rhr} Y$	$\Rightarrow$	$X \leq_{st} Y$

Table: Relationships among the aging classes introduced above. Here "$A \longrightarrow B$" stands "A implies B", where $\longleftarrow^{*}$ holds only if $-\infty < l < r = +\infty$, while $\longrightarrow^{**}$ holds only if $-\infty = l < r < +\infty$, with $l, r \in \mathbb{R}$ denoting lower and upper limits of the support.

		DRHR	$\longleftarrow$	ILR	$\longrightarrow$	IHR		
	$\nearrow$						$\nwarrow$	
DD								ID
	$\nwarrow$						$\nearrow$	
		DHR	$\longleftarrow^{*}$	DLR	$\longrightarrow^{**}$	IRHR		

Weighted or biased distributions are used to perturb an original distribution by multiplying a baseline PDF with a suitable weight function. In particular, for a given $w : \mathbb{R} \to \mathbb{R}$ non-negative weight function, if X has an absolutely continuous distribution with PDF f, then the weighted random variable X_w has PDF defined as

$$f_w(t) = \frac{w(t)f(t)}{\mu_w}, \qquad t \in \mathbb{R}, \tag{1}$$

by assuming that $\mu_w = \int_{\mathbb{R}} w(t)f(t)\mathrm{d}t$ is finite and non-zero.

Generalized equilibrium distributions

If $w(t) = 1/\tau(t)$ for $t < 0$ and $w(t) = 1/\lambda(t)$ for $t \geq 0$, then we get the following (recent) generalization of the equilibrium distribution.

Definition

Let X be a non-degenerate random variable, having CDF F, SF $\overline{F}$ and finite mean. The generalized equilibrium random variable of X, denoted as X^e, has the following PDF

$$f^e(t) = \begin{cases} \dfrac{F(t)}{\tilde{\mu}}, & t < 0, \\ \dfrac{\overline{F}(t)}{\tilde{\mu}}, & t \geq 0, \end{cases} \tag{2}$$

where $\tilde{\mu} = \mathrm{E}(|X|) > 0$.

Suitable choices for the weight w in (1) lead to novel skewing strategies of a given baseline PDF.

Definition

Let us assume that X has an absolutely continuous distribution with PDF f, while Y has CDF G and SF $\overline{G}$, respectively. If X and Y are independent, then the left-skewed random variable X_L has PDF

$$f_L(t) = \frac{\overline{G}(t)f(t)}{\Pr(Y > X)}, \qquad t \in \mathbb{R}, \tag{3}$$

while the right-skewed random variable X_R has PDF

$$f_R(t) = \frac{G(t)f(t)}{\Pr(Y \le X)}, \qquad t \in \mathbb{R}. \tag{4}$$

Note that $X_L \le_{lr} X \le_{lr} X_R$.

Definition

A distortion function is an increasing continuous function $q : [0,1] \to [0,1]$, such that $q(0) = 0$ and $q(1) = 1$.

The distorted CDF from F through q is given by $F_q = q(F)$. In addition, the dual distortion function with respect to q is defined as

$$\tilde{q}(u) = 1 - q(1-u), \qquad u \in [0,1]$$

and, thus, $\overline{F}_q = \tilde{q}(\overline{F})$ is the distorted SF of F_q.

...gracias Miguel Angel y Alfonso!

ESTADISTICA ESPAÑOLA
Vol. 33, Núm. 127, 1991, págs. 257 a 284

Deformación de funciones de distribución: una técnica estadística

por
JULIO MIRAS
Instituto Nacional de Estadística

RESUMEN

Si F(*x*) es una función de distribución y G(*y*) es otra función de distribución cuya masa está contenida en el intervalo (0,1), la composición H(*x*)=G(F(*x*)), bajo ciertas condiciones, es una nueva función de distribución. Denominamos a H(*x*) Deformación de F(*x*) engendrada por el generador G(*y*). Se propone una técnica que consiste en postular una distribución F(*x*) que representa el modelo base o inicial en un problema de estadística, y una familia de generadores {G(*y*;Θ)} de modo que H(*x*;Θ) = G(F(*x*);Θ) representa el conjunto de alternativas a la distribución base.

Palabras clave: Comparación de funciones de distribución, Modelos de inferencia estadística.

Clasificación AMS: 62E99.

ROC and OD curves

The Receiver Operating Characteristic (ROC) curve is a plot of the True Positive Rate (TPR, sensitivity or probability of detection) against the False Positive Rate (FPR, probability of false alarm or error of type I) at each threshold setting in a binary diagnosis (classification) test.

A dual approach to the ROC analysis is described by the Ordinal Dominance (OD) curve. Indeed, the OD curve is a plot of the true negative rate (specificity or selectivity) against false negative rate (miss rate or error of type II) at each threshold setting in a binary diagnosis test.

ROC and OD curves

The Receiver Operating Characteristic (ROC) curve is a plot of the True Positive Rate (TPR, sensitivity or probability of detection) against the False Positive Rate (FPR, probability of false alarm or error of type I) at each threshold setting in a binary diagnosis (classification) test.

A dual approach to the ROC analysis is described by the Ordinal Dominance (OD) curve. Indeed, the OD curve is a plot of the true negative rate (specificity or selectivity) against false negative rate (miss rate or error of type II) at each threshold setting in a binary diagnosis test.

Table of Contents

Definition

Let X and Y be random variables with absolutely continuous SFs $\overline{F}$ and $\overline{G}$, respectively. We assume that X and Y have an interval support with the same initial point. Then, the ROC distortion curve between X and Y is defined as

$$ROC_{\overline{G},\overline{F}}(u) = \overline{G}(\overline{F}^{-1}(u)), \qquad u \in [0,1]. \tag{5}$$

Definition

Let X and Y be random variables with absolutely continuous CDFs F and G, respectively. We assume that X and Y have an interval support with the same initial point. Then, the OD distortion curve between X and Y is defined as

$$OD_{G,F}(u) = G(F^{-1}(u)), \qquad u \in [0,1]. \tag{6}$$

A useful index to summarize the ROC curve is the area under the ROC (AUROC) which is defined by

$$AUROC_{\overline{G},\overline{F}} = \int_0^1 ROC_{\overline{G},\overline{F}}(u)\mathrm{d}u. \tag{7}$$

Similarly, we define the area under the OD (AUOD) as

$$AUOD_{G,F} = \int_0^1 OD_{G,F}(u)\mathrm{d}u \tag{8}$$

and clearly one has $AUROC_{\overline{G},\overline{F}} + AUOD_{G,F} = 1$.

Proposition

If X and Y are independent, then

$$AUROC_{\overline{G},\overline{F}} = \Pr(Y > X), \qquad AUOD_{G,F} = \Pr(Y \leq X).$$

Example

Let X be exponentially distributed with SF $\overline{F}(t) = e^{-\lambda t}$, for $t \geq 0$ and $\lambda > 0$. Let Y be Weibull distributed with SF $\overline{G}(t) = e^{-(\lambda t)^\alpha}$, for $t \geq 0$, $\lambda > 0$ and $\alpha > 0$. Then, from Eq. (5), we get

$$ROC_{\overline{G},\overline{F}}(u) = e^{-(-\ln(u))^\alpha}, \qquad u \in [0,1],\ \alpha > 0, \tag{9}$$

that is plotted in the left-hand side of the next figure for $\alpha = 1/3, 1/2, 1, 2$. For such choices of α one respectively has $AUROC_{\overline{G},\overline{F}} = 0.4311101, 0.4543586, 0.5, 0.5456414$. In addition, from Eq. (6), we have

$$OD_{G,F}(u) = 1 - e^{-(-\ln(1-u))^\alpha}, \qquad u \in [0,1],\ \alpha > 0, \tag{10}$$

which is plotted in the right-hand side of the next figure for $\alpha = 1/3, 1/2, 1, 2$. For such choices of α one respectively gets $AUOD_{G,F} = 0.5688899, 0.5456414, 0.5, 0.4543586$.

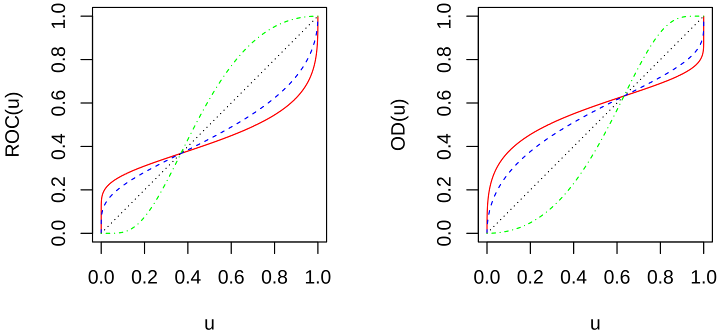

Figure: Plots of the ROC distortion curve given in Eq. (9) (left) and of the OD distortion curve given in Eq. (10) (right), for $u \in [0, 1]$ and $\alpha = 1/3, 1/2, 1, 2$ (full, dashed, dotted and dotdashed, respectively).

Connections with weighted distributions and Lorenz curve

Let X have absolutely continuous CDF F and SF $\overline{F}$. Let X_w have CDF F_w, SF $\overline{F}_w$ and PDF as in Eq. (1), according to a non-negative weight function w. Then, from Eq. (5), it holds

$$ROC_{\overline{F}_w,\overline{F}}(u) = \frac{1}{\mu_w}\int_0^u w(\overline{F}^{-1}(v))\mathrm{d}v, \qquad u \in [0,1]$$

and, from Eq. (6), one has

$$OD_{F_w,F}(u) = \frac{1}{\mu_w}\int_0^u w(F^{-1}(v))\mathrm{d}v, \qquad u \in [0,1]. \tag{11}$$

Remark

If X is non-negative and $w(t) = t$, then $\mu_w = \mu$ and we get

$$L(u) = OD_{F_w,F}(u) = \frac{1}{\mu}\int_0^u F^{-1}(v)\mathrm{d}v, \qquad u \in [0,1], \tag{12}$$

which is the Lorenz curve of X.

Connections with weighted distributions and Lorenz curve

Let X have absolutely continuous CDF F and SF $\overline{F}$. Let X_w have CDF F_w, SF $\overline{F}_w$ and PDF as in Eq. (1), according to a non-negative weight function w. Then, from Eq. (5), it holds

$$ROC_{\overline{F}_w,\overline{F}}(u) = \frac{1}{\mu_w}\int_0^u w(\overline{F}^{-1}(v))\mathrm{d}v, \qquad u \in [0,1]$$

and, from Eq. (6), one has

$$OD_{F_w,F}(u) = \frac{1}{\mu_w}\int_0^u w(F^{-1}(v))\mathrm{d}v, \qquad u \in [0,1]. \tag{11}$$

Remark

If X is non-negative and $w(t) = t$, then $\mu_w = \mu$ and we get

$$L(u) = OD_{F_w,F}(u) = \frac{1}{\mu}\int_0^u F^{-1}(v)\mathrm{d}v, \qquad u \in [0,1], \tag{12}$$

which is the Lorenz curve of X.

Table of Contents

Table: CDFs, SFs, quantile functions and means of relative random variables having support $[0,1]$. Here X and Y have CDFs F and G, SFs $\overline{F}$ and $\overline{G}$, respectively.

Random variable	$V = F(Y)$	$S = \overline{F}(Y)$	$W = G(X)$	$T = \overline{G}(X)$
CDF	$OD_{G,F}(u)$	$ROC_{\overline{G},\overline{F}}(u)$	$OD_{F,G}(u)$	$ROC_{\overline{F},\overline{G}}(u)$
SF	$ROC_{\overline{G},\overline{F}}(1-u)$	$OD_{G,F}(1-u)$	$ROC_{\overline{F},\overline{G}}(1-u)$	$OD_{F,G}(1-u)$
Quantile function	$OD_{F,G}(u)$	$ROC_{\overline{F},\overline{G}}(u)$	$OD_{G,F}(u)$	$ROC_{\overline{G},\overline{F}}(u)$
Mean	$AUROC_{\overline{G},\overline{F}}$	$AUOD_{G,F}$	$AUROC_{\overline{F},\overline{G}}$	$AUOD_{F,G}$

Proposition

Let X and Y be random variables with absolutely continuous CDFs F and G, SFs $\overline{F}$ and $\overline{G}$ and PDFs f and g, respectively. Let us denote with $V = F(Y)$ and $W = G(X)$ and let U be a random variable with a standard uniform distribution. The following statements are equivalent:

(i) $X \leq_{lr} Y$.

(ii) $U \leq_{lr} V$.

(iii) $W \leq_{lr} U$.

(iv) $OD_{G,F}(u)$ convex for $u \in [0,1]$.

(v) $ROC_{\overline{G},\overline{F}}(u)$ concave for $u \in [0,1]$.

(vi) V is ID.

Proposition

Let X and Y be random variables with absolutely continuous CDFs F and G, SFs $\overline{F}$ and $\overline{G}$ and PDFs f and g, respectively. Anyone of the following statements implies that $V = F(Y)$ is IHR:

(i) $X \leq_{lr} Y$.

(ii) X *is DHR and* Y *is IHR.*

Proposition

Let X_1, Y_1, X_2 and Y_2 be random variables. Let us assume that $V_1 = F_1(Y_1)$ and $V_2 = F_2(Y_2)$, where F_1 and F_2 are the CDFs of X_1 and X_2, respectively. If $X_2 \leq_{st} X_1$ and $Y_1 \leq_{st} Y_2$, then $V_1 \leq_{st} V_2$.

Proposition

Let X_1, Y_1, X_2 and Y_2 be random variables. Let us assume that $V_1 = F_1(Y_1)$ and $V_2 = F_2(Y_2)$, where F_1 and F_2 are the CDFs of X_1 and X_2, respectively. If $X_2 \leq_{hr} X_1$ and $Y_1 \leq_{hr} Y_2$, with X_1 DHR and Y_1 IHR, then $V_1 \leq_{hr} V_2$.

Table of Contents

Let X and Y be random variables with absolutely continuous SFs $\overline{F}$ and $\overline{G}$, respectively, having an interval support with the same initial point. By recalling Eq. (5), the partial AUROC between X and Y for a given interval (u_1, u_2) is defined by

$$pAUROC_{\overline{G},\overline{F}}(u_1, u_2) = \int_{u_1}^{u_2} ROC_{\overline{G},\overline{F}}(v)\mathrm{d}v, \qquad 0 \leq u_1 \leq u_2 \leq 1, \quad (13)$$

which represents the AUROC between two given thresholds u_1 and u_2. Moreover, from Eqs. (7) and (13), the relative AUROC between X and Y for the interval (u_1, u_2) is defined by

$$rAUROC_{\overline{G},\overline{F}}(u_1, u_2) = \frac{pAUROC_{\overline{G},\overline{F}}(u_1, u_2)}{AUROC_{\overline{G},\overline{F}}}, \qquad 0 \leq u_1 \leq u_2 \leq 1, \quad (14)$$

which can be interpreted as the proportion of the AUROC being captured between two given thresholds u_1 and u_2.

Proposition

Let X be a random variable having absolutely continuous SF $\overline{F}$. Let Y be a random variable with SF $\overline{G}$, independent from X. Then

$$rAUROC_{\overline{G},\overline{F}}(u_1, u_2) = ROC_{\overline{F}_L,\overline{F}}(u_2) - ROC_{\overline{F}_L,\overline{F}}(u_1), \qquad 0 \leq u_1 \leq u_2 \leq 1,$$

where $\overline{F}_L$ is the SF of the random variable X_L with PDF given in Eq. (3).

As a consequence of the latter proposition and $X_L \leq_{lr} X$, the following function

$$rAUROC_{\overline{G},\overline{F}}(0, u) = ROC_{\overline{F}_L,\overline{F}}(u) = \overline{F}_L(\overline{F}^{-1}(u)), \qquad u \in [0, 1], \qquad (15)$$

is a proper convex distortion function, namely the relative AUROC distortion between X and Y.

Let us denote with F and G the CDFs of X and Y, respectively. The partial AUOD between X and Y for a given interval (u_1, u_2) is defined by

$$pAUOD_{G,F}(u_1, u_2) = \int_{u_1}^{u_2} G(F^{-1}(v))\mathrm{d}v, \qquad 0 \leq u_1 \leq u_2 \leq 1, \qquad (16)$$

quantifying the AUOD between two given thresholds u_1 and u_2. Moreover, from Eqs. (8) and (16), the relative AUOD between X and Y for the interval (u_1, u_2) is defined by

$$rAUOD_{G,F}(u_1, u_2) = \frac{pAUOD_{G,F}(u_1, u_2)}{AUOD_{G,F}}, \qquad 0 \leq u_1 \leq u_2 \leq 1, \qquad (17)$$

which represents the proportion of the AUOD being captured between two given thresholds u_1 and u_2.

Proposition

Let X be a random variable having absolutely continuous CDF F. Let Y be a random variable with CDF G, independent from X. One has

$$rAUOD_{G,F}(u_1, u_2) = OD_{F_R,F}(u_2) - OD_{F_R,F}(u_1), \qquad 0 \le u_1 \le u_2 \le 1,$$

where F_R is the CDF of the random variable X_R with PDF given in Eq. (4).

From the latter proposition and $X \le_{lr} X_R$, the following function

$$rAUOD_{G,F}(0, u) = OD_{F_R,F}(u) = F_R(F^{-1}(u)), \qquad u \in [0, 1], \qquad (18)$$

is a proper convex distortion function, namely the relative AUOD distortion between X and Y.

Table: CDFs, SFs, quantile functions and means of the equilibrium version of the relative random variables with support in $[0,1]$ given in the previous table. Here X_L (X_R) denotes the left (right) skewed version of X with weight $\overline{G}$ (G), while Y_L (Y_R) denotes the left (right) skewed version of Y with weight $\overline{F}$ (F).

Random variable	$V^e =_{st} F(X_L)$	$S^e =_{st} \overline{F}(X_R)$	$W^e =_{st} G(Y_L)$	$T^e =_{st} \overline{G}(Y_R)$
CDF	$OD_{F_L,F}(u)$	$ROC_{\overline{F}_R,\overline{F}}(u)$	$OD_{G_L,G}(u)$	$ROC_{\overline{G}_R,\overline{G}}(u)$
SF	$ROC_{\overline{F}_L,\overline{F}}(1-u)$	$OD_{F_R,F}(1-u)$	$ROC_{\overline{G}_L,\overline{G}}(1-u)$	$OD_{G_R,G}(1-u)$
Quantile function	$OD_{F,F_L}(u)$	$ROC_{\overline{F},\overline{F}_R}(u)$	$OD_{G,G_L}(u)$	$ROC_{\overline{G},\overline{G}_R}(u)$
Mean	$AUROC_{\overline{F}_L,\overline{F}}$	$AUOD_{F_R,F}$	$AUROC_{\overline{G}_L,\overline{G}}$	$AUOD_{G_R,G}$

Table of Contents

Using the ideas previously discussed on ROC and OD distortion functions, we evaluate the performance of four representative Machine Learning (ML) binary classification methods, such as

- logistic regression (LR),
- support vector machine (SVM),
- multi layer perceptron (MLP) with a single hidden layer,
- random forest (RF) ensemble algorithm.

Let us denote by B the binary outcome, taking values 0 or 1 for each one of the categories. Let us denote by S a scoring diagnosis variable being determined by the outomes of a ML classifier (the measures of a biomarker or by any other measurable variable) aimed at predicting the target class $B = 1$ of the binary outcome. On the other hand, let us consider

$$X =_{st} (S|B = 0) \quad \text{and} \quad Y =_{st} (S|B = 1),$$

with CDFs F and G, SFs $\overline{F}$ and $\overline{G}$, respectively.

(i) The empirical estimation of the ROC can be calculated from the scores provided by S evaluated on sample data, that is

$$\hat{R}_{n,m}(u) = 1 - \hat{G}_n(\hat{F}_m^{-1}(1-u)), \qquad u \in [0,1],$$

where $\hat{F}_m$ and $\hat{G}_n$ are the empirical CDFs, estimating respectively F and G, with n and m denoting the respective sample sizes for the categories $B = 1$ and $B = 0$ in the sample data.

(ii) An alternative approach is concerned with a kernel-based estimation which employs a smooth kernel PDF estimation of class $B = 0$ and $B = 1$ observations to calculate the ROC curve.

(iii) We consider the PHR Cox model, described by $\overline{G}_\theta(x) = (\overline{F}(x))^\theta$ for $\theta > 0$, where $\overline{F}$ and $\overline{G}_\theta$ are the SFs of X and Y, respectively. In this case one has

$$ROC_{\overline{G}_\theta,\overline{F}}(u) = \overline{G}_\theta(\overline{F}^{-1}(u)) = u^\theta, \qquad u \in [0,1], \tag{19}$$

which is a concave distortion function for $\theta \in (0,1)$ and a convex one for $\theta > 1$. From Eqs. (7) and (19), it follows that $AUROC = 1/(1+\theta)$, from which a semiparametric estimation of the ROC can be calculated as

$$\hat{R}_{n,m}(u;\hat{\theta}_{PHR}) = u^{\hat{\theta}_{PHR}}, \qquad u \in [0,1], \tag{20}$$

where $\hat{\theta}_{PHR} = -1 + 1/MW$ is the estimated value of θ, with MW denoting the Mann-Whitney non-parametric estimation of the AUROC.

(iv) Another option concerns with the PO model, described by

$$\frac{\overline{G}_\theta(x)}{G_\theta(x)} = \theta\frac{\overline{F}(x)}{F(x)}, \quad \theta > 0.$$

Therefore, under the PO model, one has

$$ROC_{\overline{G}_\theta,\overline{F}}(u) = \overline{G}_\theta(\overline{F}^{-1}(u)) = \frac{\theta u}{1+(\theta-1)u}, \qquad u \in [0,1], \qquad (21)$$

which is a concave distortion when $\theta > 1$ and a convex one for $\theta \in (0,1)$. From Eqs. (7) and (21), it turns out that

$$AUROC_{\overline{G}_\theta,\overline{F}} = \int_0^1 \frac{\theta u}{1+(\theta-1)u}\mathrm{d}u = \frac{\theta^2-\theta-\theta\log\theta}{(\theta-1)^2}, \qquad (22)$$

which leads to the following semiparametric estimation of the ROC

$$\hat{R}_{n,m}(u;\hat{\theta}_{MO}) = \frac{\hat{\theta}_{MO}\cdot u}{1+(\hat{\theta}_{MO}-1)u}, \qquad u \in [0,1], \qquad (23)$$

where $\hat{\theta}_{MO}$ is the estimated value of θ.

All the previous mentioned approaches are applied to assess the performance of LR, SVM, MLP and RF classifiers to predict customer churn in a Telco company. The data set consists of a sample of 3333 customers for whom several input business variables have been collected along with a binary outcome taking the value $B = 1$ if the customer has left the company (churn event) and the value $B = 0$ otherwise.

Table: Input variables collected and measured in the churn data set.

Variable	Description
VMailMessage	Number of voice mail messages
DayMins	Service use during the day (in minutes)
DayCalls	Total number of calls during the day
EveMins	Service use during the evening (in minutes)
EveCalls	Total number of calls during the evening
NightMins	Service use during the night (in minutes)
NightCalls	Total number of calls during the night
IntlMins	Service use for international calls (in minutes)
IntlCalls	Total number of international calls
CustServCalls	Number of calls to the customer service

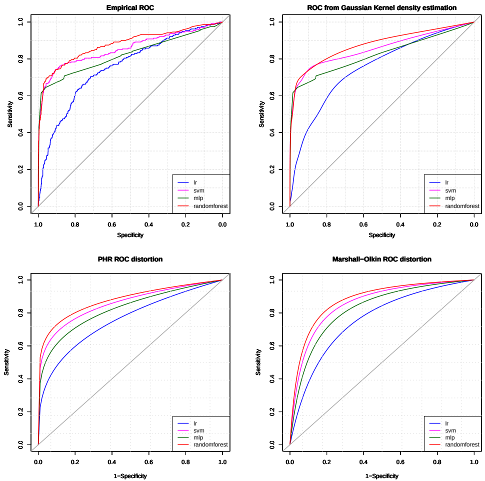

Figure: Plots for ROC comparisons applied to the assessment of LR, SMV, MLP and RF classifiers for churn prediction: Empirical ROC (top left), ROC from Gaussian kernel PDF estimation (top right), PHR ROC distortion (bottom left) and Marshall-Olkin ROC distortion (bottom right).

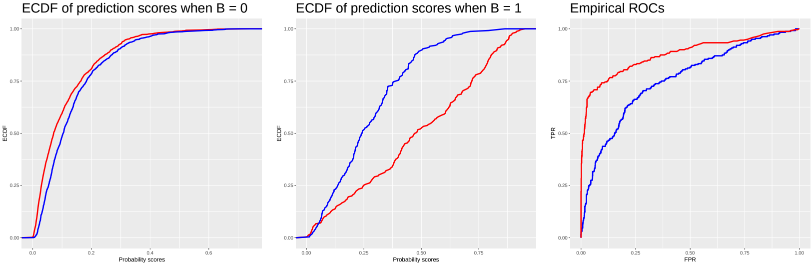

Figure: Empirical CDF comparisons of probability scores for $B = 0$ (left) and $B = 1$ (middle) as well the corresponding empirical ROC comparison of LR (blue) and RF (red) churn predictors (right).

Table of Contents

Final remarks

(i) We have studied the ROC distortion curve. Analogously, the OD curve has been examined as a distortion too. This fact allows us to explicate the distortion function which connects two SFs (or, respectively, CDFs) when distortion-based model assumptions are unknown.

(ii) It is also helpful to interpret the Lorenz curve as a distortion function which results from the OD distortion between a non-negative random variable and its length biased version.

(iii) Several stochastic comparisons and aging results have been provided as well, by taking into account the interpretation of ROC and OD distortions as CDFs of suitable relative random variables.

(iv) Relative AUROC distortion and relative AUOD distortion have been defined. Both findings have close connections with recent skewing mechanisms and with the concept of equilibrium distribution.

(v) Our theoretical findings are illustrated in different settings through real data applications of the distortion-based approach by using a semiparametric estimation of the ROC curve.

Final remarks

(i) We have studied the ROC distortion curve. Analogously, the OD curve has been examined as a distortion too. This fact allows us to explicate the distortion function which connects two SFs (or, respectively, CDFs) when distortion-based model assumptions are unknown.

(ii) It is also helpful to interpret the Lorenz curve as a distortion function which results from the OD distortion between a non-negative random variable and its length biased version.

(iii) Several stochastic comparisons and aging results have been provided as well, by taking into account the interpretation of ROC and OD distortions as CDFs of suitable relative random variables.

(iv) Relative AUROC distortion and relative AUOD distortion have been defined. Both findings have close connections with recent skewing mechanisms and with the concept of equilibrium distribution.

(v) Our theoretical findings are illustrated in different settings through real data applications of the distortion-based approach by using a semiparametric estimation of the ROC curve.

Final remarks

(i) We have studied the ROC distortion curve. Analogously, the OD curve has been examined as a distortion too. This fact allows us to explicate the distortion function which connects two SFs (or, respectively, CDFs) when distortion-based model assumptions are unknown.

(ii) It is also helpful to interpret the Lorenz curve as a distortion function which results from the OD distortion between a non-negative random variable and its length biased version.

(iii) Several stochastic comparisons and aging results have been provided as well, by taking into account the interpretation of ROC and OD distortions as CDFs of suitable relative random variables.

(iv) Relative AUROC distortion and relative AUOD distortion have been defined. Both findings have close connections with recent skewing mechanisms and with the concept of equilibrium distribution.

(v) Our theoretical findings are illustrated in different settings through real data applications of the distortion-based approach by using a semiparametric estimation of the ROC curve.

Final remarks

(i) We have studied the ROC distortion curve. Analogously, the OD curve has been examined as a distortion too. This fact allows us to explicate the distortion function which connects two SFs (or, respectively, CDFs) when distortion-based model assumptions are unknown.

(ii) It is also helpful to interpret the Lorenz curve as a distortion function which results from the OD distortion between a non-negative random variable and its length biased version.

(iii) Several stochastic comparisons and aging results have been provided as well, by taking into account the interpretation of ROC and OD distortions as CDFs of suitable relative random variables.

(iv) Relative AUROC distortion and relative AUOD distortion have been defined. Both findings have close connections with recent skewing mechanisms and with the concept of equilibrium distribution.

(v) Our theoretical findings are illustrated in different settings through real data applications of the distortion-based approach by using a semiparametric estimation of the ROC curve.

Final remarks

(i) We have studied the ROC distortion curve. Analogously, the OD curve has been examined as a distortion too. This fact allows us to explicate the distortion function which connects two SFs (or, respectively, CDFs) when distortion-based model assumptions are unknown.

(ii) It is also helpful to interpret the Lorenz curve as a distortion function which results from the OD distortion between a non-negative random variable and its length biased version.

(iii) Several stochastic comparisons and aging results have been provided as well, by taking into account the interpretation of ROC and OD distortions as CDFs of suitable relative random variables.

(iv) Relative AUROC distortion and relative AUOD distortion have been defined. Both findings have close connections with recent skewing mechanisms and with the concept of equilibrium distribution.

(v) Our theoretical findings are illustrated in different settings through real data applications of the distortion-based approach by using a semiparametric estimation of the ROC curve.

Some references

1. Bamber D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology **12**, 387–415.
2. Calì C., Longobardi M. (2015). Some mathematical properties of the ROC curve and their applications. Ricerche di Matematica **64**, 391–402.
3. Capaldo M., Arevalillo J.M., Navarro J. (2025). Distorted distributions and ROC curves. Scandinavian Journal of Statistics **52**, 1786–1815.
4. Capaldo M., Di Crescenzo A., Navarro J. (2025). Generalized equilibrium distributions. Probability in the Engineering and Informational Sciences, 1–27. Online first.
5. Cox D.R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34** 187–202.
6. Marshall A.W., Olkin I. (1997) A new method for adding a parameter to a family of distribution with application to the exponential and Weibull families. Biometrika **84**, 641–652.

Some references

7. Navarro J. (2022). Introduction to System Reliability Theory. Springer.
8. Navarro J., Arevalillo J.M. (2023). On connections between skewed, weighted and distorted distributions: applications to model extreme value distributions. Test **32**, 1307–1335.
9. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **12**, 77.
10. Shaked M., Shanthikumar J.G. (2007). Stochastic Orders. Springer, New York.
11. Wang S. (1996). Premium calculation by transforming the layer premium density. ASTIN Bulletin: The Journal of the IAA **26**, 71–92.
12. Yaari M.E. (1987). The dual theory of choice under risk. Econometrica: Journal of the Econometric Society **85**, 95–115.

Grazie per la vostra attenzione!