Statistics and Probability Letters 103 (2015) 73–79

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On shape properties of the receiver operating characteristic curve

Bhaskar Bhattacharya [a,*], Gareth Hughes [b]

[a] Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA

[b] Scotland's Rural College, The King's Buildings, West Mains Road, Edinburgh, EH9 3JG, UK

ARTICLE INFO

Article history:
Received 3 August 2014
Received in revised form 31 March 2015
Accepted 1 April 2015
Available online 23 April 2015

Keywords:
Asymmetry
Kullback–Leibler divergence
Entropy
Relative distributions
ROC

ABSTRACT

We present formal definitions of two commonly observed asymmetries in a concave receiver operating characteristic curve. The main theorem of the paper proves that the Kullback–Leibler divergences between the underlying signal and noise variables are ordered based on these asymmetries. This result is true for *any* continuous distributions of the signal and noise variables.

1. Introduction

For binary diagnostic decisions, the receiver operating characteristic (ROC) curve is a method for assessing the performance of procedures to classify subjects (or a received piece of information) as *signal* or *noise*. Imagine a single threshold on the scale of an underlying indicator variable. Then the group of subjects with indicator scores above the threshold denotes a prediction of need for treatment and the group of subjects with indicator scores below the threshold denotes a prediction of no need for treatment. Signals that are correctly classified are 'true positives', and noises that are correctly classified are 'true negatives'. Then the true positive proportion (TPP) expresses the number of true positives as a proportion of the total number of signals and the true negative proportion (TNP) expresses the number of true negatives as a proportion of the total number of noises. The false positive proportion is $FPP = 1 - TNP$. Pairs of TPP and FPP values are obtained by allowing the threshold indicator score to vary over the whole range of the indicator variable. An ROC curve is a graphical plot of TPP against FPP (Krzanowski and Hand, 2009).

In this paper, we consider two absolutely continuous random variables (rvs), X for signal and Y for noise, with common support. Let X have cumulative distribution function (cdf) F_1 and probability density function (pdf) f_1, and, those for Y are F_2 and f_2, respectively. For any cdf F, we define the inverse function $F^{-1}(r) = \sup\{x : F(x) \leq r\}, 0 \leq r \leq 1$. Assume F_1, F_2 are invertible. It is well known that $ROC(t) = S_1(S_2^{-1}(t)), 0 \leq t \leq 1$ (Krzanowski and Hand, 2009), where $S_i(x) = 1 - F_i(x), i = 1, 2$. Then $ROC(t)$ is invertible. We assume that both ROC and ROC^{-1} are differentiable, and $ROC(t)$ is concave in t.

Using X and Y, we define two new rvs, known as the *relative* rvs. Consider the transformation $U = F_2(X)$, which takes values in $(0, 1)$; U is the relative rv which denotes the relative distribution for comparing X with reference to Y

* Corresponding author.
E-mail address: bhaskar@siu.edu (B. Bhattacharya).

http://dx.doi.org/10.1016/j.spl.2015.04.003
0167-7152/

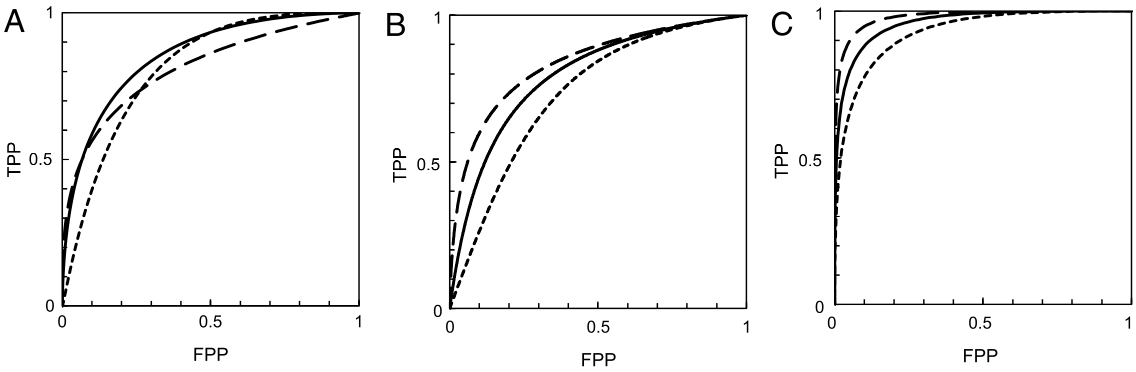

Fig. 1. ROC curves with varying parameters (Table 1). Solid line is symmetric ROC; dashed line is TPP-asymmetric ROC; dotted line is TNP-asymmetric ROC. A. Binormal ROC curves. B. Lloyd's (Lloyd, 2000) two-parameter lomax model ROC curves. C. Bi-beta ROC curves (Marzban, 2004).

(Handcock and Morris, 1998). It is well known that the cdf and the pdf of U are

$$G_1(u) = F_1(F_2^{-1}(u)), \qquad g_1(u) = \frac{f_1(F_2^{-1}(u))}{f_2(F_2^{-1}(u))}, \quad 0 \le u \le 1, \tag{1}$$

respectively. If the distributions of X and Y are identical, then the cdf of U is a 45° line and the pdf of U is the uniform (0, 1) distribution.

Similarly, suppose $V = F_1(Y)$ is the relative rv which denotes the relative distribution for comparing Y with reference to X, then the cdf and the pdf of V are $G_2(v) = F_2F_1^{-1}(v)$, $g_2(v) = \frac{f_2(F_1^{-1}(v))}{f_1(F_1^{-1}(v))}$, $0 \le v \le 1$, respectively. Here G_1, G_2 are invertible, as F_1, F_2 are so.

For continuous signal and noise populations with pdfs $f_1(x)$, $f_2(x)$, $x \in \mathcal{D} \subset \Re$, respectively, the *Kullback–Leibler divergence* (KLD) between $f_1(x)$, $f_2(x)$ is defined as $I(f_1|f_2) = \int f_1(x) \ln\left(\frac{f_1(x)}{f_2(x)}\right) dx$. Although $I(f_1|f_2)$ is not a metric, it is well known that $I(f_1|f_2) \ge 0$ and $I(f_1|f_2) = 0$ if and only if $f_1(x) = f_2(x), \forall x \in \mathcal{D}$. Hence it is often interpreted as a measure of 'divergence' or 'distance' between f_1 and f_2. Here and in the sequel, we observe the conventions that $\ln 0 = -\infty$, $\ln(a/0) = +\infty$ $(a > 0)$, $0 \cdot (\pm\infty) = 0$. Interchanging the roles of f_1 and f_2, one could define $I(f_2|f_1)$ in a similar manner. However, it is well known that, in general, $I(f_1|f_2) \ne I(f_2|f_1)$ (Cover and Thomas, 1991).

Entropy (or differential entropy) of an rv X with cdf $F(x)$ (absolutely continuous) and pdf $f(x)$ is defined (and expressed) as $H(X) = -\int \ln f(x) dF(x) = -\int f(x) \ln f(x) dx$. Entropy is a measure of disparity of a density function ($f(x)$) from the uniform distribution. It measures the uncertainty of X in the sense of utility of using $f(x)$ in place of the ultimate uncertainty of the uniform distribution (Good, 1968). For discrete rvs, entropy is nonnegative. However for a continuous rv X, entropy takes values in $[-\infty, \infty]$.

2. Examples and motivation

To see how changes in the shape of an *ROC* curve correspond to changes in KLD, consider Fig. 1 and Table 1. In Fig. 1, each of the three boxes (unit squares), has three *ROC* curves drawn in it. The area inside the box is referred to as the *ROC space*. Now imagine the line joining points (0, 1) and (1, 0) of the unit square, known as the *negative diagonal*. When the *ROC* curve is symmetric about the negative diagonal, it is called a *symmetric ROC*, otherwise it is called an *asymmetric ROC*. Each box in Fig. 1 (A, B or C) has one symmetric (solid line) and two asymmetric (dashed or dotted line) *ROC* curves in it. The underlying signal and noise distributions (normal, lomax or beta) giving rise to these *ROC* curves are listed in Table 1.

As in Section 1, recall f_1: signal, f_2: noise. Also, Table 1 lists the KLDs $[I(f_1|f_2)$ and $I(f_2|f_1)]$ corresponding to the distributions f_1, f_2. In each case (A, B or C), we see from Table 1 that for symmetric *ROC*, the KLDs are equal (measured in a unit called 'nits').

To describe asymmetric *ROC* curves, we consider *true positive* or *negative proportion* (TPP or TNP) types; TPP-asymmetry happens when the *ROC* is more inclined (or, 'skewed' e.g. in Taylor, 1967) towards the TPP axis (y-axis in *ROC* space) when compared with a symmetric *ROC*, and TNP-asymmetry happens when the *ROC* is more inclined towards the top line in the *ROC* space (whose equation is $y = 1$ in Cartesian plane), often called the TNP axis, when compared with a symmetric *ROC*, as seen from Fig. 1. For asymmetric *ROC*s, when we compare the KLDs from Table 1, we find: for TPP-asymmetry, $I(f_1|f_2) > I(f_2|f_1)$ holds, and for TNP-asymmetry, $I(f_1|f_2) < I(f_2|f_1)$ holds. These observations are validated in Theorem 5.

Researchers are motivated to study the symmetry and asymmetry properties of the *ROC* curves because it leads to information about the underlying (unknown) signal and noise variables. An *ROC* curve is called *binormal* if both signal and noise variables are normally distributed, say $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, respectively. Discussions in Green and Swets (1966) considered a binormal symmetric *ROC* when $\mu_1 > \mu_2$ with $\sigma_1^2 = \sigma_2^2$, and an asymmetric *ROC*, when $\mu_1 > \mu_2$ with $\sigma_1^2 > \sigma_2^2$.