[Alfred et. al., Vol.6 (Iss.1): January, 2018]
(Received: Jan 13, 2018 - Accepted: Jan 30, 2018)
ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.29121/granthaalayah.v6.i1.2018.1652

INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH
A knowledge Repository

Science

COMPARISION OF BINARY DIAGNOSTIC PREDICTORS USING ENTROPY

Kathare Alfred [1], Otieno Argwings [2], Kimeli Victor [3]
[1, 2, 3] Department of Mathematics and Computer Sciences, University of Eldoret

Abstract

The use of gold standard procedures in screening may be costly, risky or even unethical. It is, therefore, not admissible for large scale application. In this case, a more acceptable diagnostic predictor is applied to a sample of subjects alongside a gold standard procedure. The performance of the predictor is then evaluated using Receiver Operating Characteristic curve. The area under the curve, then, provides a summative measure of the performance of the predictor. The Receiver Operating Characteristic curve is a trade-off between sensitivity and specificity which in most cases are of different clinical significance. Also, the area under the curve is criticized for lack of coherent interpretation. In this study, we proposed the use of entropy as a summary index measure of uncertainty to compare diagnostic predictors. Noting that a diseased subject who is truly identified with the disease at a lower cut-off will also be identified at a higher cut-off, we substituted time variable in survival analysis for cut-offs in a binary predictor. We then derived the entropy of the functions of diagnostic predictors. Application of the procedure to real data showed that entropy was a strong measure for quantifying the amount of uncertainty engulfed in a set of cut-offs of binary diagnostic predictor.

Keywords: Entropy; Binary Diagnostic; Predictors.

Cite This Article: Kathare Alfred, Otieno Argwings, and Kimeli Victor. (2018). "COMPARISION OF BINARY DIAGNOSTIC PREDICTORS USING ENTROPY." *International Journal of Research - Granthaalayah*, 6(1), 440-447. https://doi.org/10.29121/granthaalayah.v6.i1.2018.1652.

1. Introduction

Screening among the population at risk of a disease has historically been important element of disease control. The use of error-free ("gold" standard) test during screening may potentially pose high risk, expensive or sometimes unethical for large scale application. In this case other acceptable diagnostic binary diagnostic predictor that sorts the subjects either as having the disease (diseased) or not (disease-free) is first applied across a range set of cut-off points against gold standard to a selected sample. Its performance is then evaluated before large scale application. For example, if mammography was used to screen for breast cancer, the gold standard would be pathological classification of tissue biopsy. At any one cut-off point, the

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

diagnostic predictor yields either a true positive (TP), false positive (FP), false negative (FN) or a true negative (TN) result.

Traditionally, the performance of a binary diagnostic predictor has been evaluated using receiver operating characteristic (ROC) curve with the area under the curve (AUC) providing a summative measure of its performance. The ROC curve is simply a graph of true positive rate (sensitivity) against false positive rate (I –specificity). Two important issues arise from this approach. One, sensitivity (proportion of test-positive among subjects who have the disease) and specificity (proportion of test-negative among subjects who do not have the disease) have different clinical consequences. Thus, depending on clinical importance of the test, one may be interested in either sensitivity or specificity and not the trade-off between the two. For instance, where treatment is expensive or involves potentially high risk procedures, a diagnostic predictor with high level of sensitivity may be preferred as opposed to low false positive rate (1 – specificity). The other issue is that the area under the ROC curve lacks clinical interpretation.

Where two or more binary predictive diagnostic predictors are available, one would like to quantitatively compare their sensitivities and select one diagnostic predictor over the other to roll out a screening program. In this paper, we describe how one would derive a descriptive summary measure, *entropy*, of the functions of a binary diagnostic predictor. The new index is an easy to compute. Its interpretation is based on information theory as the amount of uncertainty that the diagnostic predictor is able to provide within a set of consecutive cut-off points. Using a real data, we demonstrate how the index can be used to compare two or more diagnostic predictors.

2. Methodology

2.1. Study Design

A simple randomized screening test design was used for the study. In this design, the study individuals are first classified by gold standard as “diseased” or “non-diseased”. These individuals are then randomly assigned for screening to one of the two or more binary classifiers over a set of ordered cut-offs.

Let C denote the outcome of a binary predictive classifier and G be the outcome of the gold standard such that

$$C = \begin{cases} 1 \; if \;\; the \; test \; result \; is \;\; positive \\ \\ 0 \; if \;\; the \; test \; results \; is \; negative \end{cases} \quad \text{and} \quad G = \begin{cases} 1 \; if \;\; a \; subject \; is \; diseased \\ \\ 0 \; if \;\; a \; subject \; is \; non-diseased \end{cases}$$

Also let $i = 1, 2, ..., k$ denote a particular value of the random variable X representing the cut-off point and $q = 1, 2, ..., s$ be the particular predictive classifier.

The quantity $n_{cg}^{i\,q}$ represents results of the gold standard classification and screening by the predictive classifier where c and g represents the binary classifier and gold standard respectively. Then, the ordered tables of these values for a fixed q form cumulative partial tables for given reference cells. Thus, for true positive we have $n_{11}^{k\,q} > \mathrm{n}_{11}^{(k-1)\,q} > \ldots > \mathrm{n}_{11}^{2\,q} > n_{11}^{1\,q}$.

Essentially, this means that if a diseased case was correctly identified at cut-off X_i it will also be correctly identified at cut-off X_{i+1}. Similarly for false negative we have $n_{01}^{1q} > n_{01}^{2q} > \ldots > n_{01}^{(k-1)q} > n_{01}^{kq}$.

For false positive we have $n_{10}^{kq} > n_{10}^{(k-1)q} > \ldots > n_{10}^{2q} > n_{10}^{1q}$ and for true negative we have $n_{00}^{1q} > n_{00}^{2q} > \ldots > n_{00}^{(k-1)q} > n_{00}^{kq}$.

2.2. Derivation of Entropy of the Sensitivity of a Classifier

In this section, we apply the concept of survival analysis to derive the entropy of true positive values. In survival analysis, the objects are observed over non-negative random variable T representing the waiting time until the event occurs. Here we substitute the time variant with cut-off of predictor. Thus, we observe the N subjects over some ordered cut-offs.

The confirmed diseased subjects n will be distributed cumulatively across non-negative random variable X, $\alpha < x < \varpi$ representing the cut-offs. Using n as the radix we can define the probability of "survivors" (those not yet identified yet they have the disease) across X. Let $f(x)$ be the probability density function and $F(x)$ be the cumulative distribution function of X respectively.

By definition $F(x) = \Pr\left(X \le x\right)$ ……………………………………………………… 2.1

$F(x)$ gives the cumulative probability that the subject was correctly identified by cut-off x. The function $F(x)$ is the true positive rate at x such that $\mathrm{F}(\varpi) = 1$ and $\mathrm{F}(\alpha) = 0$.

The survival function is be given by

$S(x) = \Pr\left\{X > x\right\} = 1 - F(x) = 1 - \int_{\alpha}^{\varpi} f(x)dx$ ……………………………………………... 2.2

The survival function gives the probability that a diseased subject is not yet correctly identified by cut-off x. This means in the lowest cut-off α, none of the diseased subject is identified with the disease while in the highest cut-off ϖ all the diseased cases will be identified as diseased.

$$\hat{S}(x) = \frac{Number\ of\ diseased\ cases\ not\ yet\ identified\ by\ cut-off\ x}{confirmed\ number\ of\ diseased\ cases} \qquad \ldots\ldots\ldots\ldots\ldots 2.3$$

Equation 2.3 represents the false negatives at cut-off x.

Let $\lambda(x)$ be some hazard function representing the instantaneous rate of correctly identifying the diseased subjects within some criteria interval dx of the predictor. By definition

$$\lambda(x) = \lim_{dx \to 0} \frac{\Pr\left\{x < X \le x + dx / X > x\right\}}{dx} \ldots\ldots\ldots\ldots 2.4$$

The numerator of the hazard function can therefore be written as the ratio of the joint probability that X is in the interval $x, x+dx$ and $X > x$ to the probability of the condition $X > x$. Thus

$$\lambda(x) = \lim_{dx \to 0} \frac{\Pr\left\{x < X \le x + dx\ and\ X > x\right\}}{\Pr(X > x)dx} \ldots\ldots\ldots\ldots 2.5$$

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

But $\Pr\{x < X \le x + dx \text{ and } X > x\} = f(x)dx$ for small dx and $\Pr(X > x) = S(x)$

Hence $\lambda(x) = \lim_{dx \to 0} \frac{\Pr\{x < X \le x + dx \text{ and } X > x\}}{\Pr(X > x)dx} = \frac{f(x)dx}{S(x)dx} = \frac{f(x)}{S(x)}$ ……..…… 2.6

This implies $\lambda(x) = \frac{f(x)}{S(x)}$ is the hazard function.

Now $\frac{d}{dx}S(x) = \frac{d}{dx}\{1 - F(x)\} = -\frac{d}{dx}F(x) = -f(x) \Rightarrow -\frac{d}{dx}S(x) = f(x)$ …….....…. 2.7

This means $\lambda(x) = \frac{f(x)}{S(x)} = \frac{-\frac{d}{dx}S(x)}{S(x)} = -\frac{d}{dx}\ln S(x)$ ……..…...….... 2.8

Thus $\lambda(x) = -\frac{d}{dx}\ln S(x)$ ….……..…..……...……..... 2.9

Equation 2.9 represents the proportion of change of sensitivity over interval dx.
The survival function (false negative function) can be expressed as a function of hazard function (true positive function) over some cut-offs such that for $S(x_0) = 1$ we have

Thus $\int_{x_0}^{x} \lambda(x)dx = -\int_{x_0}^{x} \frac{\frac{d}{dx}S(x)}{S(x)}dx = -\int_{x_0}^{x} \frac{d}{dx}\ln S(x)dx = -\ln S(x)\Big|_{x_0}^{x} = -\ln S(x)$ ……..…. 2.10

It is to be noted that $S(x_0) = 1$ and thus $\ln(1) = 0$.

$\int_{x_0}^{x} \lambda(x)dx = -\ln S(x) \quad \Leftrightarrow \quad S(x) = \exp\left\{-\int_{x_0}^{x} \lambda(x)dx\right\}$ ……..……...……..…… 2.11

Thus $S(x) = \exp\left(-\int_{x_0}^{x} \lambda(x)dx\right) = e^{-\psi(x)}$ ……...……...…….. 2.12

In this case, we have $S(x) = e^{-\psi(x)}$ where $\psi(x) = \int_{x_0}^{x} \lambda(x)dx$ is the cumulative hazard function of X.

Now the expectation of X is given by $E[X] = \mu$. By definition

$\mu^0 = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx = -\int_{0}^{\infty} x\frac{d}{dx}S(x)dx$ ……...…….. 2.13

$\mu^0 = -\int_{0}^{\infty} x\frac{d}{dx}S(x)dx$. Equation 2.13 represent the mean cut-off point.

Integrating $\int_{0}^{\infty} x\frac{d}{dx}S(x)dx$ by parts $\int udv = uv - \int vdu$

we have $u^0 = -\int_{0}^{\infty} x\frac{d}{dx}S(x)dx = -\left[xS(x) - \int S(x)dx\right]_0^{\infty} = -xS(x)\Big|_{x=0}^{x \to \infty} + \int_{0}^{\infty} S(x)dx$

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

Now $S(0)=1$ and $S(\infty)=0$. Then $\mu^0=\int_0^\infty S(x)dx$ ……..…….…. 2.14

Putting it in simple terms, the mean of X is the integral of the survival function $S(x)$. The survival function $S(x)$ in our case is the false negative function.

It is then possible to link the expectation of X and the hazard function. Thus

$\mu^0=\int_0^\infty S(x)dx=\int_0^\infty e^{-\varphi(x)}dx$ ……..……...…..…….. 3.15

Suppose now that $\psi(x)$ is changed by $100\delta\%$ to become $\psi(x)(1+\delta)$. We show how this translates into changes in expectation of X. First $S(x)$ becomes $S^*(x)$.

$S^*(x)=\exp\left(-\int_0^\infty \psi(x)(1+\delta)\right)dx=S(x)^{1+\delta}$ ……..……. 2.16

$S^*(x)=S(x)^{1+\delta}$

Then the new expectation of the false negative function becomes

$\mu^*=\int_0^\infty S^*(x)dx=\int_0^\infty S(x)^{1+\delta}$ ……..……...……...…… 2.17

$\mu^*=\int_0^\infty S(x)^{1+\delta}$

To find the effect of δ on the expectation of X we find the derivative of μ^* with respect to δ. In this case

$\frac{d\mu^*}{d\delta}=\int_0^\infty \ln S(x).S(x)^{1+\delta}dx$ ……...……..... 2.18

and within the neighborhood of $\delta=0$ we get

$\frac{d\mu^*}{d\delta}=\int_0^\infty \ln S(x).S(x)dx$ ……...……...……...…….....…….. 2.19

Now $\frac{\Delta\mu^*}{\Delta\delta}=\int_0^\infty \ln S(x).S(x)dx$ implying $\Delta\mu^*=\Delta\delta\int_0^\infty \ln S(x).S(x)dx$

$\frac{\Delta\mu^*}{\mu}=\frac{\Delta\delta\int_0^\infty \ln S(x).S(x)dx}{\int_0^\infty S(x)dx}=-H(X)\Delta\delta$ ……...……...……...……..... 2.20

The quantity $H(X)=-\frac{\int \ln S(x).S(x)dx}{\int S(x)dx}$ ……...……...……..... 2.21

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

is the entropy of the survival function (false negatives). In general, the entropy of any probability density function $f(x)$ is given by $H(X) = -\frac{-\int \ln F(x).F(x)dx}{\int F(x)dx} = \frac{-\sum \ln F(x).F(x)}{\sum F(x)}$. It is similar to Shannon entropy $H(X) = E\left[-\log_b \mathrm{P(X)}\right] = -\sum_i p(x_i)\log_b p(x_i)$ except that $\mathrm{F}(x)$ is a cumulative distribution function of X and that the former is weighted with $\frac{1}{\int \mathrm{F}(x)dx}$

For $b = 2$ the unit of entropy is *bit*, for $b = e$ the unit of entropy is *nat* and for $b = 10$ the entropy unit is *dit* (or *digit*).
In the event $p(x_i) = 0$ for some i the value of the corresponding summand $0\log_b 0$ is taken to be 0 which is consistent with $\lim_{p\to 0+} p\log(p) = 0$.

Entropy is a measure of the uncertainty in a random variable (Martin et al, 2011). It can be as low as zero ($H(X) = 0$) if the sensitivity of the classifier was 100% for all values of X. In this case, the ROC curve would be a straight line on the upper side from $[0,1]$ to $[1,1]$. The highest value of $H(X)$ can be got from the criteria defining a horizontal line from $[0,0]$ to $[1,1]$. The closer H is to zero the better the sensitivity of the predictor. When $H(X) = 0$, the classifier is perfect and taking the subjects through the entire criteria provides no information.

Just like partial area under the ROC curve, partial entropy can be computed for a set of cut-offs if known to be of clinical importance. In this case we integrate with the interval of under consideration.

3. Application of Entropy to Real Data

We compared sensitivities of two classifiers measuring the carbohydrate antigen 19-9 (CA 19-9). Elevated levels of CA 19-9 (> 37 U/mL) has been found to be associated with gastrointestinal carcinomas particularly in pancreatic cancer. We thus, bench marked our cut-off point at 40 U/m and weighted cut-off above 40 U/Ml nearly twice to spread the possibility of cancer detection. Our cut-offs thus ranged between X >110 and X >0 at arbitrarily interval of 10 U/mL. Entropies of the true positive functions of both diagnostic predictors were estimated as $H_1\ (X)_{se}$ 0.3079 and $H_2\ (X)_{se}$ 0.7535. The results of entropy show that diagnostic predictor 2 delivered more than twice the information delivered by diagnostic predictor 1. Overall, thus, diagnostic predictor 1 was twice likely to correctly identify a subject with a disease compared to diagnostic predictor 2.

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

4. Conclusion

Entropy enjoys variety of interpretation and therefore used in a wide range of disciplines to measure the degree of randomness in a system. Using a simple screening design, we substituted time variable in survival analysis for cut-offs in binary diagnostic predictor and demonstrated how disorder in a binary predictor can be assessed using entropy of any of its four functions; true positive rate, false positive rate, false negative rate and true negative rate. Depending on the clinical importance of these functions, their entropies can be used to compare the amount of uncertainty that the diagnostic predictors derives across a set of cut-off or criteria.

References

[1] Adams, N. M., & Hand, D. J. (1999). Comparing diagnostic predictors when the misallocation costs are uncertain. *Pattern Recognition*, *32*, 1139–1147.
[2] Akobeng, A.K. (2006). Understating diagnostic tests 1: Sensitivity, specificity and predictive values. *Act Pediatric*, 96, 338-341.doi:10.1111/j.1651-2227. 2006. 00180.x.
[3] Bradley, P. A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *The Pennsylvania State University*, *30*(7), 1145--1159. doi: 10.1016/ S0031-3203(96)00142-2.x.
[4] Cheng, H., & Macaluso, M. (1996). Comparison of the accuracy of two tests with a confirmatory procedure limited to positive results. *Epidemiology*, *8*, 104–106.
[5] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. doi:10.1016/j.patrec.2005.10.010.x
[6] Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49.
[7] Griner, P., Mayewski, R. J., Mushlin, A. I., & Greenlan, P. (1981). Selection and interpretation of diagnostic tests and procedures: Principles and applications. *Annals of Internal Medicine*, *94*, 557–592.
[8] Halligan, S., Douglas, G. A., & Mallet, S.(2015). Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *European Radiology*, 25(4), 932–939. doi: 10.1007/s00330-014-3487-0. x.
[9] Hanley, A. J. (1989). Receiver Operating Characteristics (ROC) Methodology the State of the Art.
[10] *Critical Reviews in Diagnostic Imaging*, 29(3), 307-335. doi: 10.1007/s00330-014-3487-*0.x.*
[11] Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36. doi: 10.1148/7063747.x
[12] Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, *48*(4), 277–287.
[13] Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151. doi: 10.1111/j.1466-8238.2007.00358.x.
[14] Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction* (1st ed.). United Kingdom: Oxford University Press.
[15] Pepe, M.S. (2000) Receiver operating characteristic methodology. *Journal of the American Statistical Association*, 95, 308–311.
[16] Rothman, K. J. (1986). *Modern Epidemiology*. Boston; Brown and Company.
[17] Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American*, 283(4), 82–87.
[18] Thompson, M. L., & Zucchini, W. (1989). On the statistical analysis of ROC curves. *Statistics in Medicine*. 8(10), 1277-1290.

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1169035

[19] Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistical Medicine*, 24(13), 2025–2540. doi: 10.1002/sim.2103.x.

[20] Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*. 39(4), 561–577.