A Note on the Area under the Gains Chart

Chris J. Lloyd*

University of Melbourne, Carlton, 3053, Australia

Abstract: The Receiver Operating Characteristic (ROC) chart is well known in medicine and machine learning. In particular the area under the ROC chart measures the probability of correct selection in a two alternative forced choice (2AFC) scenario. The gains chart is closely related to the ROC curve but carries extra information about the rate at which the classifier identifies response, information that is not carried by the ROC chart. In this note, we point out that the appropriate area under the gains chart is identical to the analogous area under the ROC chart and that the gains chart is therefor to be preferred as a summary of classifier success.

Keywords: ROC curve, gains chart, 2AFC.

1. INTRODUCTION

Binary classification is a standard problem in medical science as well as the general machine learning sphere. One of the standard measures of the success of the classifier is a plot of the true positive fraction (TPF) against the false positive fraction (FPF). In statistical parlance, this is a plot of the power of the test versus the size as the cut-off is varied from one extreme to the other. In business applications, an alternative but closely related curve is used, known as the *gains chart*. This plots the true positive fraction against the unconditional positive fraction. The purpose of this chart is to track how fast responders are accumulated as we work our way down the list from highest ranked to lowest ranked based on our model.

The two charts look quite similar except the slope of the ROC chart is vertical at the beginning whereas the slope of the gains chart gives direct information about the rate of response accumulation. The area under the ROC chart has a specific interpretation, first identified by Bamber (1975) [1], namely that it estimates the probability of success in a 2AFC experiment. Indeed, estimation of this easily interpreted area under the ROC curve is a key reason for its popularity. In this note, we point out that the area under the gains chart can be similarly interpreted (with a minor modification).

2. MATHEMATICAL DEFINITIONS

Suppose we have a binary classifier that produces a continuous diagnostic x, larger values of which indicate response. Without loss of generality, we assume that x is scaled to the unit interval and may think of it as a fitted probability of response, for instance based on a logistic regression or a neural network. There are two charts that measure the rate at which large values of the diagnostic accumulate responders. The ROC chart plots this against the rate at which we accumulate errors and the gain chart against the rate at which we make assessments.

Let $F_J(x)$ denote the distribution function of x for responders (J=1) and non-responders (J=0). The unconditional distribution F of x is

$$F(x) = \pi_1 F_1(x) + (1 - \pi_1)F_0(x)$$

where π_1=Pr(response).

We will assume x is continuous though there is no essential difference in results if x is discrete once one accounts for the occurrence of ties in the definition of the curves. Basically, once the "steps" in the curves can be joined by lines the continuous theory applies.

2.1. ROC Chart

The ROC curve measures how the probability of a true and false positive covary when we classify individuals as responders if x > γ. Specifically it is a plot of

$$y = 1 - F_1(\gamma) \text{ versus } x = 1 - F_0(\gamma)$$

as we vary γ from 0 to 1. It is a simple matter to express γ in terms of x to obtain the explicit form

$$R(x) = 1 - F_1(F_0^{-1}((1 - x)).$$

A perfect classifier, where the distribution of x for the responders and non-responders are completely separated, will increase from 0 to 1 as the cut-off moves past the largest value of x for the non-responders. The ROC curve increases from 0 to 1 at

*Address correspondence to this author at the University of Melbourne, Carlton, 3053, Australia; Tel: 613 93498228;
E-mail: c.lloyd@mbs.edu

E-ISSN: 1929-6029/18

the origin. A classifier that has no predictive power will follow the diagonal, $R(x) = x$. A typical example is in the left panel of Figure **1** based on a classifier of fake bank-notes based on 610 fake bank-notes and 762 genuine bank-notes. The green point represents the performance of the classifier when bank-notes are classified as fake when the estimated probability is greater than 0.5. Apparently, we identify around 18% of genuine notes as fake and 78% of fake bank notes as fake.

There is a well-developed theory on estimating this curve from the different kinds of data that may be available. Estimation under a normal model from ordinal data was solved by Dorfman & Alf (1968) [2] but F_0 and F_1 must both be transformable to normal by the same transformation. There are many examples where the bi-normal model is not justified. Non-parametric and semi-parametric theory was established by Hsieh & Turnbull (1996) [3]. Smoothing methods were first considered by Lloyd (1998) [4] and shown to be superior to fully empirical in Lloyd and Zhou (1999) [5]. Nevertheless, fully empirical estimators based on the empirical estimators of F_1 and F_0 are most commonly used, as displayed in the example in the Figure.

The area under this curve

$$\text{AUR} = \int_0^1 R(x)dx = \int_0^1 (1 - F_1(F_0^{-1}(1-x)))dx$$

is a common summary of overall accuracy. It was proven by Bamber (1975) [1] that this area measures the probability that two randomly selected individuals from the response and non-response population will be correctly classified by the classifier, known as 2AFC. This is closely related to the Mann-Whitney U-statistic for testing the difference between two populations, here the responders and non-responders with respect to the diagnostic x. Since a completely random classifier will have AUR of 0.5 and a perfect classifier 1.0, it is natural to rescale this as 2 × AUR − 1. This is the area below the ROC but above the random diagonal as a proportion of the area above the diagonal.

In the left panel of Figure **1**, the area under the curve is 0.875 which means that if we are presented with two bank-notes known to be randomly chosen, one from the fake and one from the genuine population, then we will correctly identify which is which 87.5% of the time. The rescaled version of this statistics is 37.5%/50%=75%. So, this classifier is three quarters of the way between a random classifier and a perfect classifier.

2.2. Gains Chart

The *cumulative gains chart* (often just called the gains chart) is a plot of the true positive probability against the unconditional positive probability i.e.

$$y = 1 - F_1(\gamma) \text{ versus } x = 1 - F(\gamma).$$

In a medical testing context, it shows how quickly we identify target patients as we work our way through a population of patients who have been rated by our classifier from most likely to least likely. Again, for a random classifier this plot will just be the diagonal. For a perfect classifier, it will increased from 0 to 1 as the proportion of overall positives increases from 0 to π_1, because all of these are true positives.

The right panel of Figure **1** displays the gains chart for the same bank-notes classifier as for the ROC chart to the left. The proportion of fake notes was 44.4%.

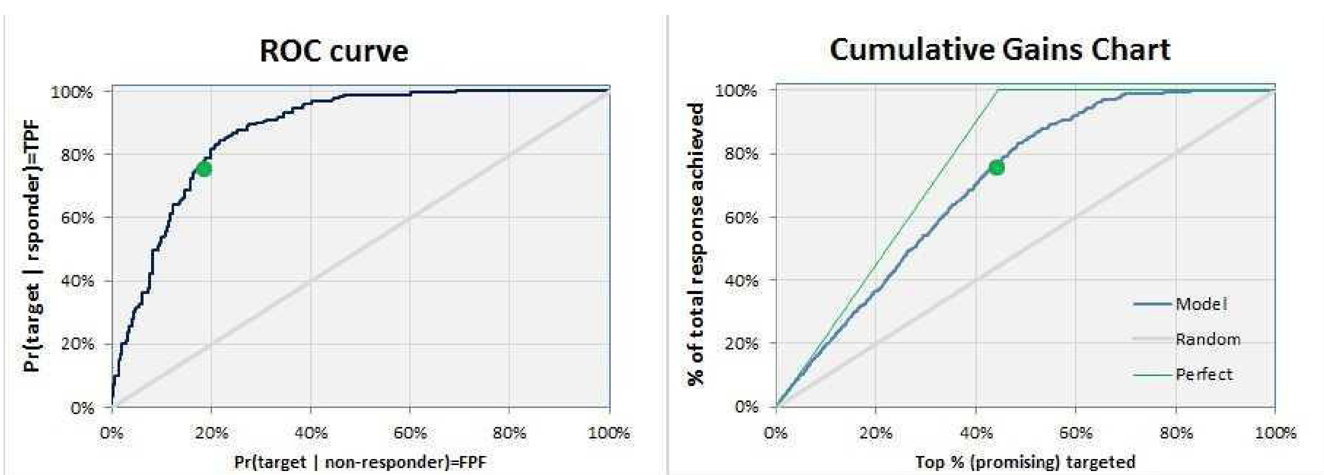

Figure 1: Summary graphs for a classifier. Left. ROC curve. Right. Gains chart.

The green point represents the classifier based on a 50-50 cut-off. This targets 43% of the bank-notes while correctly identifying 78% of the fake bank-notes. The gains charts can be calibrated against the perfect classifier, whose gains chart is displayed in green, where we identify the notes as fake without error.

The similarity of the shapes of the two curves is obvious and a consequence of plotting the same quantity against a different coordinate.

Following the practice for ROC curves, it would be natural to measure the overall accuracy of the gains chart by the area beneath. But this can never exceed a maximum possible area of $1-0.5\pi_1$ for the perfect classifier. Since the area between the perfect chart and the diagonal is $0.5(1-\pi_1)$, an analogous measure to $2 \times \text{AUR} - 1$ would be

$$\frac{\text{AUG}-0.5}{0.5(1-\pi_1)} = \frac{2\text{AUG}-1}{1-\pi_1}$$

3. MAIN RESULT

It is convenient in what follows to express the inverse of the ROC curve as

$$R^{-1}(x) = 1 - F_0(F_1^{-1}(1-x))$$

obtained by reversing the response and non-response subscripts. Note that the area beneath this curve is 1 − AUR.

The inverse G^{-1} of the gains function would be a plot of $x=1-F(\gamma)$ against $y=1-F_1(\gamma)$ as γ varies. Substituting $\gamma = F_1^{-1}(1-y)$ we find that

$$G^{-1}(y) = x = \pi_1((1 - F_1(F_1^{-1}(1-y))) + (1-\pi_1)(1 - F_0(F_1^{-1}(1-y))$$

Hence

$$G^{-1}(y) = \pi_1 y + (1-\pi_1)R^{-1}(y).$$

The inverse of the gains function is a probability weighted average of the gains function y of the random classifier and the inverse of the ROC function. Integrating both sides

$$1 - \text{AUG} = 0.5\pi_1 + (1-\pi_1)(1-\text{AUR})$$

from which it easily follows that

$$\frac{\text{AUG}-0.5}{0.5(1-\pi_1)} = 2\,\text{AUR}-1.$$

This is the main result which equates the rescaled area under the ROC curve with the rescaled area under the gains chart.

Referring to the right panel of Figure **1**, the area beneath the gains chart is 0.708. The area between the perfect and random classifier is 0.5×(1−0.444) = 0.278. So the area between the gains chart and the diagonal (0.208) as a proportion of the difference between a perfect and random classifier (0.278) is 0.208/ 0.278=0.75, which is the value we obtained from the ROC chart.

4. CONCLUSION

We have shown what was perhaps already well-known in the folklore, namely that the gains chart is a simple re-expression of the ROC chart. Indeed, the inverse gains chart is a linear combination of the inverse ROC curve and the random diagonal. The standard measure of how much better a classifier is than random, namely 2AUR−1 from the ROC curve, has a numerically identical analogue for the gains chart. We conclude that the gains chart has all the advantages of the ROC curve but also provides information on the overall proportion of responses (π_1) and the rate at which we identify these responses as the classifier searches a population ranked by the classifier score x.

Other summary measures of ROC besides area beneath have been considered, see for instance Wieand *et al.* (1989) [6]. These will all no doubt have obvious analogues for the gains chart.

REFERENCES

[1] Bamber D. The area above the ordinal dominance graph and the area below the receiver operating graph. Journal of Mathematical Psychology 1975; 12: 387. https://doi.org/10.1016/0022-2496(75)90001-2

[2] Dorfmann DD, Alf AE. Maximum likelihood estimation of parameters of signal detection theory - a direct solution. Psychometrika 1968; 33: 113. https://doi.org/10.1007/BF02289677

[3] Hsieh F, Turnbull BW. Non-parametric and semi-parametric estimation of the receiver operating characteristic curve. Ann Statistics 1996; 24: 25. https://doi.org/10.1214/aos/1033066197

[4] Lloyd CJ. The use of smoothed ROC curves to summarise and compare diagnostic systems. Journal of the American Statistical Association 1998; 93: 1356. https://doi.org/10.1080/01621459.1998.10473797

[5] Lloyd CJ, Zhou Y. Kernel estimators of the ROC curve are better than empirical. Statistics and Probability Letters 1999; 44: 221.
https://doi.org/10.1016/S0167-7152(99)00012-7

[6] Wieand S, Gail M, James B, James K. A family of non-parametric statistics for comparing diagnostic markers with paired and unpaired data. Biometrika 1989; 76: 585.
https://doi.org/10.1093/biomet/76.3.585

Received on 16-05-2018 Accepted on 04-06-2018 Published on 25-06-2018

https://doi.org/10.6000/1929-6029.2018.07.03.1