Improved Methods for Estimating Areas under the Receiver Operating Characteristic Curves: A Confidence Interval Approach

R. Vishnu Vardhan, Department of Statistics, Pondicherry University, Kalapet, Puducherry, India
S. Balaswamy, Department of Statistics, Pondicherry University, Kalapet, Puducherry, India

ABSTRACT

ROC Curve is the most widely used statistical technique for classifying an individual into one of the two pre-determined groups basing on test result. Area under the curve (AUC) is a measure of accuracy which exhibits the discriminating power of the test with respect to a threshold or cutoff value. In medical diagnosis, this technique has its relevance to study and compare different diagnostic tests. In this paper, a method is proposed to estimate the AUC of Binormal ROC model by taking into account the confidence interval of mean and corresponding variances.

Keywords: Accuracy, Area Under the Curve (AUC), Binomal Receiver Operating Characteristic (ROC), Confidence Interval of Mean, Receiver Operating Characteristic (ROC) Curve

INTRODUCTION

In the recent years, the Receiver Operating Characteristic (ROC) curve analysis has become a popular statistical technique in the field of medical diagnosis. Green and Swets reported that the ROC curve was originated during Second World War, later it made its landmark in diversified fields such as Medicine, Experimental Psychology, Finance, Banking, Data Mining etc.

Apart from well-known statistical classification procedures like Logistic Regression and Discriminant analysis, the ROC curve has also its mathematical formulation which helps in fitting and estimating the parameters of the

DOI: 10.4018/jgc.2013070105
curve. The entire classification is carried out on the basis of a threshold value often referred to as Gold Standard, it determines the true condition status and also provides a source of information completely different from the tests under evaluation. If the condition status is true, it indicates the presence of disease otherwise absence. Two basic measures of ROC curves are sensitivity ($S_n$) and specificity ($S_p$).

Sensitivity refers to the ability of a test to detect the condition when it is present and Specificity refers to the ability of test to exclude patients without the condition.

An ROC curve is a plot of $1-S_p$ versus $S_n$ (Krazonowski & Hand, 2009). The construction of ROC curve primarily depends on the four possible states which are obtained on the basis of a threshold value i. e., TP, TN, FN and FP. The resulting curve is called empirical ROC. Bamber (1975) stated that diseased (X) and the healthy (Y) individuals follow Normal distribution and hence the name Binormal ROC curve, with unknown monotonic transformation.

There are two main objectives of ROC curve analysis. The first one is to identify the best optimal cutoff in some sense and the other one is to choose among several procedures (called biomarkers) the best one in terms of AUC.

**LITERATURE REVIEW**

In parametric approach, much work has been done over the years. This approach constitutes robust mathematical structures, where these help in evaluating the performance of a diagnostic test and also enables to assess the effect of covariates in the model. We reviewed some of the recent contributions in this approach. It is noticed that, during a decade of time, many researchers have contributed much in developing the ROC curve model by incorporating the covariate effects, different estimating procedures etc.

James A. Hanley (1988) has proved the robustness of using the bi-normal model to fit the ROC curves. Pepe M. S. (2000) has proposed a General linear model approach for evaluating the ROC form and as well as estimating the AUC. Zang Zheng, Pepe, M. S. (2005) has been introduced a linear regression frame work to evaluate the ROC model by involving the covariate effects on the ROC curves. Fawcett Tom et. al, (2007) studied the non-convex ROC curves and the conditions that can lead to empirical and fitted ROC curves that are not convex.

Metz, C. E. (2010) has proposed a proper bi-normal ROC curve for improving the accuracy. The terminology used for proper bi-normal ROC curve is the ‘hook’. This means that the new defined ROC curve will get boosted at the curvilinear part and provides better accuracy when compared to the conventional form of Binormal ROC curve. Ethan J, et al. (2010) have projected the Comparison of ROC curves on the basis of optimal operating point. Davidov, O., & Nov, Y. (2011) proposed an improved, consistent and asymptotically normally distributed estimator which outperforms the original estimator of Hsieh and Turnbull (1996).

Halpern, Schisterman, Enrique F et. al. (2010) estimated the ROC curve and gave an approximation for the confidence interval of the linear combination of multiple bio-markers, by assuming that the covariates are adjusted and follow multivariate normal distribution.

Farragi and Reiser (2002) have proved that the bi-normal model performs very similar to the non-parametric approach by simulation studies.

Andriy, I. B., Howard, E. R., Tao, S. & David, G. (2009) have proposed a new index of the free-response performance at all decision thresholds simultaneously and this can be regarded as an area under the empirical free response ROC curve. They have also developed a non-parametric method for its analysis and also derived the ideal bootstrap estimator of the variance.

Huang, Y., & Pepe, M. S. (2009) have developed the methods for making inference about the predictiveness curve using cross-sectional or cohort data. They have investigated the relationship between the ROC curve and the predictiveness curve and they have provided a way of combining information across populations that have similar ROC curves but varying prevalence of the outcome.
Related Content

Evaluating the Possibilities of Improving the Quality of Tourism Services of the “Eco- Guesthouses” from Mrginimea Sibiului

The Effect of Business Characteristics on the Methods of Knowledge Protections
[www.igi-global.com/article/effect-business-characteristics-methods-knowledge/58343?camid=4v1a](www.igi-global.com/article/effect-business-characteristics-methods-knowledge/58343?camid=4v1a)
Design of Solar System for LTE Networks
www.igi-global.com/article/design-of-solar-system-for-lte-networks/258050?camid=4v1a

Pricing Rental Tourist Accommodation: Airbnb in Barcelona
www.igi-global.com/chapter/pricing-rental-tourist-accommodation/235837?camid=4v1a