12-1-2014

Overview of Inference about ROC Curve in Medical Diagnosis

Jingjing Yin
Georgia Southern University, jyin@georgiasouthern.edu

Recommended Citation
Yin, Jingjing. 2014. "Overview of Inference about ROC Curve in Medical Diagnosis." *Biometrics and Biostatistics International Journal*, 1 (3): 00013: MedCrave Group. doi: 10.15406/bbij.2014.01.00013 source: https://doi.org/10.15406/bbij.2014.01.00013
https://digitalcommons.georgiasouthern.edu/biostat-facpubs/89

Overview of Inference about Roc Curve in Medical Diagnosis

Editorial

Volume 1 Issue 3 - 2014

Jingjing Yin*

Department of Biostatistics, Georgia Southern University, USA

***Corresponding author:** Jingjing Yin, Department of Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Hendricks Hall 1007, P.O. Box 8015, Statesboro, GA 30460-8015, USA, Tel: 912-478-2413; Email: jyin@georgiasouthern.edu

Received: November 29, 2014 | **Published:** December 01, 2014

Medical diagnosis aims to identify diseased individuals through the evaluation of the measurements of some biomarkers by performing a diagnostic test based on some biomarker measurements. Biomarkers are measured on either discrete or continuous scale and continuous biomarkers are utilized more often in medical practice. This article introduces the most popular tool for evaluating continuous biomarkers: the Receiver Operating Characteristic (*ROC*) curve.

For diagnostic tests with binary disease status, each subject is categorized as either healthy or diseased. A perfectly accurate diagnostic test would identify all truly diseased individuals as diseased and healthy individuals as non-diseased. However, such scenarios rarely happen since mostly the diseased and healthy population distributions overlap. There are two types of diagnostic errors: false negative (FN) which happens when classifying a diseased individual as healthy and false positive (FP) which happens when classifying a healthy individual as diseased. The case correctly identifying a diseased subject as diseased is called true positive (TP) and the case correctly identifying a healthy subject as non-diseased is called true negative (TN). The proportion/rate of true positives (TPR) is commonly referred as "sensitivity" and the proportion/rate of true negatives (TNR) as "specificity". Sensitivity and specificity characterize the diagnostic accuracy under diseased and healthy population, respectively.

In order to construct a diagnostic test based on continuous biomarkers for binary disease status, a diagnostic threshold is needed. At the pre-specified diagnostic threshold value, paired values of sensitivity and specificity are computed to evaluate the test performance. As the threshold value decreases, sensitivity increases while specificity decreases. Therefore, a compromise between sensitivity and specificity is necessary to assess the test discriminatory accuracy. One popular way to evaluate the test performance over all possible threshold values is done by a graphical summary of the diagnostic accuracy, i.e. by plotting the pair of (1-specificity, sensitivity) for all possible threshold values to form a curve. This curve is known as the Receiver Operating Characteristic (*ROC*) curve. The *ROC* curve and its associated summary statistics are very useful in diagnostic field for the purpose of evaluating the discriminatory ability of biomarkers/diagnostic tests with continuous measurements. Extensive statistical research has been done in this field. There are reviews of statistical methods involving *ROC* curves [1-4].

There are two types of expressions for *ROC* curve: a point set or a curve. The *ROC* curve can be viewed as a point set of sensitivity and false positive rate given a diagnostic threshold value. Alternatively the *ROC* curve can be revised as a curve function of given values of false positive rate (i.e. 1-specificity). Generally, the second expression is used more often and it is equivalent as regarding sensitivity as a function of 1-specificity/false positive rate. Therefore, the confidence interval (*CI*) for the *ROC* curve is the same as *CI* of sensitivity at a given value of specificity [5-9]. Other situations require making inference on the whole *ROC* curve or partial *ROC* curve, i.e., most cases is more concerned with a range of high specificity (e.g. 80% to 95%). Likewise, it is also of interest to construct the confidence band (*CB*) for a portion of the *ROC* curve given a range of specificity or for the whole *ROC* curve [10-15]. The *CI* of *ROC* curve are different from *CB* as *CI* gives a likely interval range of sensitivity given a fixed value of specificity, while *CB* gives a curvy strip area that covers the whole *ROC* curve or partial *ROC* curve given a range of specificity, which maintains the type I error rate simultaneously for all values of specificity in the given range.

When considering the *ROC* curve as a point set of sensitivity and specificity and a value of diagnostic threshold is given or estimated, we can also construct the confidence region (*CR*) of sensitivity and specificity [16-17]. There might be some confusion between the *CR* and *CI* of the *ROC* curve: the *CI* of the *ROC* curve gives an interval range of possible values of sensitivity at a fixed value of specificity, while *CR* of (sensitivity, specificity) given a diagnostic threshold defines an elliptical area which is likely to cover the true values of (sensitivity, specificity). Similarly, an analogue of the *CB* for the *ROC* curve based on the *CR* of (sensitivity, specificity) would be a tube-like volume linking an infinite numbers of elliptical areas together, which maintain a specified type I error rate simultaneously for a given range of threshold values. Hence, for making inference about the whole or partial *ROC* curve, a confidence volume around the sample *ROC* curve is an alternative to the *CB* of the *ROC* curve.

References

1. Pepe MS (2004) The statistical evaluation of medical tests for classification and prediction. Oxford University Press, USA.
2. Shapiro DE (1999) The interpretation of diagnostic tests. Stat Methods Med Res 8(2): 113-134.

3. Zhou X-H, McClish DK, Obuchowski NA (2009) Statistical methods in diagnostic medicine, Volume 569, Wiley-Interscience.

4. Zou KH, Liu A, Bandos AI, Ohno-Machado L, Rockette HE (2011) Statistical evaluation of diagnostic performance: Topics in ROC analysis. CRC Press.

5. Hall P, Hyndman RJ, Fan Y (2004) Nonparametric confidence intervals for receiver operating characteristic curves. Biometrika 91(3): 743-750.

6. Linnet K (1987) Comparison of quantitative diagnostic tests: type I error, power, and sample size. Stat Med 6(2): 147-158.

7. Platt RW, Hanley JA, Yang H (2000) Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. Stat Med 19(3): 313-322.

8. Su H, Qin Y, Liang H (2009) Empirical likelihood-based confidence interval of roc curves. Stat Biopharm Res 1(4): 407-414.

9. Zhou XH, Qin G (2005) Improved confidence intervals for the sensitivity at a fixed level of specificity of a continuous-scale diagnostic test. Stat Med 24(3): 465-477.

10. Campbell G (1994) Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. Stat Med 13(5-7): 499-508.

11. Demidenko E (2012) Confidence intervals and bands for the binormal roc curve revisited. J Appl Stat 39(1): 67-79.

12. Horvath L, Horvath Z, Zhou W (2008) Confidence bands for roc curves. Journal of Statistical Planning and Inference 138(6): 1894-1904.

13. Jensen K, Muller HH, Schafer H (2000) Regional confidence bands for roc curves. Stat Med 19(4): 493-509.

14. Ma G, Hall W (1993) Confidence bands for receiver operating characteristic curves. Med Decis Making 13(3): 191-197.

15. Macskassy SA, Provost F, Rosset S (2005) ROC confidence bands: an empirical evaluation. In: Proceedings of the 22nd International Conference on Machine Learning. pp. 537-544.

16. Adimari G, Chiogna M (2010) Simple nonparametric confidence regions for the e-valuation of continuous-scale diagnostic tests. Int J Biostat 6(1): 1557-4679.

17. Yin J, Tian L (2014) Joint inference about sensitivity and specificity at the optimal cut-off point associated with youden index. Computational Statistics & Data Analysis 77: 1-13.