

ON ASSESSING EFFICIENCY OF AN ALTERNATIVE CLASSIFIER THROUGH SENSITIVITY AND SPECIFICITY

O. O. M. Sanni; E. T. Jolayemi; Ikoba, N. A. and Adeniyi, O. I.

Department of Statistics
University of Ilorin
Ilorin
(e-mail : omsanni@yahoo.com)

ABSTRACT

We examine the efficiency of a competing classifier through sensitivity and specificity, utilizing a Monte Carlo Study. We observed that when sensitivity or specificity or both are low, the efficiency of such classifier is poor and not desirable. We found that even with large sample size empirical efficiency does not show any appreciable difference. Our results suggest that estimation of efficiency is not good when we have small sample sizes (≤ 30). We found that if the sensitivity or specificity or both are high (≥ 0.75), such classifier have good efficiency. This is slightly more relaxed than the results by other researchers where sensitivity and specificity of .80 or higher was recommended.

1. Introduction

A variety of Health, Biological and Social Science data come in the form of cross-classified table of counts commonly referred to as contingency tables. Categorical data usually consists of counts rather than measurements resulting mostly from surveys that can be expressed only in qualitative or categorical terms. The units of a sampled population in such circumstances are cross-classified according to each of two categories such as Male/Female, Positive/Negative, Yes/No, Agree/Disagree, Alive/Dead, Present/Absent, etc. These categorical terms may also evolve from classifying aggregating quantitative variables. There might be the need to screen population for deadly diseases such as HIV/AIDS, Tuberculosis, Cholera, etc. In screening of subjects there are possibility of two errors, depending on the instrument used:

- (i) A diseased subject might be misclassified as being free of the disease.
- (ii) A subject free from the disease might be wrongly classified as having the disease.

To this effect there is a need to ensure that the above two errors are reduced to the barest minimum if they can not be completely avoided. Because there is no perfection in life, there is a need to examine property of instruments that are used for screening purposes. Thus a good screening test must be sensitive enough to be able to identify correctly all the affected subjects and also specific enough, that is, classify correctly all subjects that are free of the disease. Therefore efficiency of screening instruments is a function of its sensitivity and specificity.

Keywords : Sensitivity, Specificity, Efficiency, Competing Classifier, Screening Test.

Many researches have been conducted to examine classifiers either through each of sensitivity and specificity alone. For instance Tang et al (2002) established equivalence of tests using sensitivity. So are other researchers like McNell et al (1975), Nann (1997), Dunnett and Gent (1977), Farrington and Manning (1900), Gould (1993), Yanagawa et al (1994), Nann (1995), Chen et al (2000), Jolayemi (2007) who looked at the non-inferiority of tests using a combination of sensitivity and specificity but did not look at the equivalence of the tests through efficiency which should be an alternative goal and this is what this work looks at.

In section 2 we examine sensitivity and specificity and how they combine to provide an acceptable level of efficiency of a test instrument against a standard test. Section 3 is devoted to a Monte Carlo study of sensitivity, specificity and efficiency. Section 4 discusses the results and the study is concluded in section 5 with an example.

We illustrate sensitivity, specificity and efficiency of a test against a standard test when classification is dichotomous on the result of an hypothetical competing test against a standard test.

True situation as shown in Table 2.1 below.

Table 2.1 : Observed Competing Test Results And Its Probability Structure With Sample Size n Fixed

Competing Test	Standard Test +ve -ve	Total		Competing Test	Standard Test +ve -ve	Total
+ve	$n_{11} \ n_{12}$	n_1		+ve	$\pi_{11} \ \pi_{12}$	π_1
-ve	$n_{21} \ n_{22}$	n_2		-ve	$\pi_{21} \ \pi_{22}$	π_2
total	$n_1 \ n_2$	n		total	$\pi_1 \ \pi_2$	1

Clearly a full multinomial sampling scheme in which the sample size n is fixed in advance is ideal for this situation.

2. Parameter Estimates

It has been shown that

$$E(n_{ij}) = n\hat{\pi}_{ij} \text{ where } \hat{\pi}_{ij} = n_{ij}/n$$

$$Var(n_{ij}) = n\pi_{ij}(1 - \pi_{ij})$$

and $Cov(n_{ij}, n'_{ij}) = -n\pi_{ij}\pi'_{ij}$

See for example Agresti (1990), Sanni (1997), Lawal (2003) among others.

Sensitivity (θ_1) defined as the proportion of diseased subjects that are correctly classified as having the disease by the competing test and is given by

$$\theta_1 = \frac{\pi_{11}}{\pi_{.1}} \text{ whose estimate is } \hat{\theta}_1 = n_{11}/n_{.1}$$

Specificity (θ_2) on the other hand, defined as the proportion of subjects free of the disease that are correctly captured as not having the disease is given by

$$\theta_2 = \pi_{22}/\pi_{.2}, \text{ so that its estimate is given by } \hat{\theta}_2 = n_{22}/n_{.2}$$

While Efficiency (λ) is defined as the proportion of all subjects that are correctly classified among the population under consideration:

The efficiency, λ is given by

$$\begin{aligned} \lambda &= \pi_{11} + \pi_{22} \\ &= \frac{\pi_{11}}{\pi_{.1}} \pi_{.1} + \frac{\pi_{22}}{\pi_{.2}} \pi_{.2} \\ &= \theta_1 \pi_{.1} + \theta_2 \pi_{.2} \end{aligned}$$

so that its estimate

$$\hat{\lambda} = \frac{n_{11} + n_{22}}{n} = \theta_1 \frac{n_{.1}}{n} + \theta_2 \frac{n_{.2}}{n} \text{ where } n_{.1} + n_{.2} = n_{..} = n$$

Clearly an efficient screening test instrument is one with

$\lambda \rightarrow 1$, so that in Table 2. $\pi_{ij} \rightarrow 0$ for $i \neq j$ or $\pi_{12} = \pi_{21} \rightarrow 0$

This situation arises only if all the diseased subjects and all those free from the diseased subjects are correctly classified.

3. Monte Carlo Study

For each given configuration, the pivot probability π_{11} is fixed and 1000 simulations were generated for varying sample sizes ranging from 10 to 100. For each simulated 2×2 contingency table the efficiency and standard deviation of the efficiency are computed against the actual and are reported Tables 3.1 and 3.2. In doing this, there is no loss of generality if the margins are fixed so that the hypergeometric distribution can be imposed on the outcomes in Table 2.1, see Birch (1964) which automatically determines the other remaining cell probabilities under the fixed marginal constraints and reduces the number of possibilities that can provide the configuration shown in Table 2.1. Tables 3.1 and 3.2 show the Monte Carlo study results comparing the combination of sensitivity and specificity with efficiency as demonstrated in section 2.

Table 3.1 : Monte Carlo Study Of Efficiency Through Sensitivity And Specificity

Configuration With Fixed π_{11}		Sample Size (n)									
		10	20	30	40	50	60	70	80	90	100
$.4, .6; .7, .3; \pi_{11} = .10$ $\theta_1 = .15, \theta_2 = 0; \lambda = .10$	$\hat{\lambda}$ STD.DEV	.13 .07	.11 .08	.11 .05	.10 .07	.10 .06	.10 .06	.09 .07	.09 .06	.09 .06	.09 .06
$.4, .6; .7, .3; \pi_{11} = .20$ $\theta_1 = .50, \theta_2 = .17; \lambda = .30$	$\hat{\lambda}$ STD.DEV	.29 .12	.51 .11	.51 .11	.50 .09	.50 .09	.49 .09	.49 .09	.49 .09	.49 .08	.50 .08
$.4, .6; .7, .3; \pi_{11} = .30$ $\theta_1 = .75, \theta_2 = .33; \lambda = .50$	$\hat{\lambda}$ STD.DEV	.52 .15	.51 .11	.50 .09	.50 .10	.50 .09	.49 .09	.49 .09	.49 .09	.49 .08	.50 .08
$.4, .6; .5, .5; \pi_{11} = .30$ $\theta_1 = .75, \theta_2 = .67; \lambda = .70$	$\hat{\lambda}$ STD.DEV	.70 .04	.67 .06	.68 .04	.66 .05	.67 .04	.67 .04	.66 .05	.66 .04	.66 .04	.67 .04
$.55, .45; .6, .4; \pi_{11} = .45$ $\theta_1 = .75, \theta_2 = .75; \lambda = .75$	$\hat{\lambda}$ STD.DEV	.78 .080	.75 .06	.74 .09	.74 .05	.72 .22	.74 .06	.73 .08	.74 .05	.73 .07	.74 .05
$.7, .3; .7, .3; \pi_{11} = .40$ $\theta_1 = .57, \theta_2 = 0; \lambda = .40$	$\hat{\lambda}$ STD.DEV	.31 .14	.33 .14	.38 .15	.38 .14	.36 .12	.35 .12	.36 .12	.38 .13	.38 .12	.37 .11
$.7, .3; .7, .3; \pi_{11} = .45$ $\theta_1 = .64, \theta_2 = .17; \lambda = .50$	$\hat{\lambda}$ STD.DEV	.45 .21	.44 .14	.44 .19	.45 .12	.45 .17	.46 .11	.46 .16	.46 .11	.47 .15	.47 .10

Table3.2 : Monte Carlo Study Of Efficiency Through Sensitivity And Specificity

Configuration With fixed π_{11}		Sample Size (n)									
		10	20	30	40	50	60	70	80	90	100
$.5, .5; .5, .5; \pi_{11} = .40$ $\theta_1 = .80, \theta_2 = .80; \lambda = .80$	$\hat{\lambda}$ STD.DEV	.81 .04	.81 .04	.77 .05	.78 .05	.79 .04	.77 .05	.78 .04	.77 .05	.77 .04	.78 .04
$.5, .5; .5, .5; \pi_{11} = .45$ $\theta_1 = .90, \theta_2 = .90; \lambda = .90$	$\hat{\lambda}$ STD.DEV	1.0 .03	.91 .02	.90 .04	.90 .03	.89 .04	.89 .04	.88 .04	.90 .02	.88 .04	.89 .04
$.5, .5; .5, .5; \pi_{11} = .475$ $\theta_1 = .95, \theta_2 = .95; \lambda = .95$	$\hat{\lambda}$ STD.DEV	1.0 .03	1.0 .03	.92 .06	.96 .01	.95 .04	.96 .03	.93 .05	.95 .09	.95 .03	.95 .03
$.5, .5; .5, .5; \pi_{11} = .48$ $\theta_1 = .96, \theta_2 = .96; \lambda = .96$	$\hat{\lambda}$ STD.DEV	1.0 .32	1.0 .03	.97 .05	.95 .05	.97 .01	.97 .03	.97 .01	.96 .04	.94 .04	.96 .03
$.5, .5; .5, .5; \pi_{11} = .49$ $\theta_1 = .98, \theta_2 = .98; \lambda = .98$	$\hat{\lambda}$	1.0 .03	1.0 .05	.97 .05	.95 .05	.97 .01	.97 .03	.97 .01	.96 .04	.94 .04	.96 .03

θ_1 = Sensitivity, θ_2 = Specificity and λ = Expected Efficiency

4. Results

We observed from the results that when sensitivity or specificity or both are low, the efficiency is poor as expected. We found that even with increasing sample size the empirical efficiency does not show any appreciable difference.

We examined a configuration with varying combinations of sensitivity and specificity for varied sample sizes. The efficiencies as reported in Tables 3.1 and 3.2 indicated that the efficiency is not a function of the sample size. Although for sample sizes below 30, efficiencies are over estimated in most cases while in a few cases under estimation of efficiency are noted. This suggests that estimation of efficiency is not good when sizes are small (≤ 30).

When both sensitivity and specificity are at least .75 the efficiency obtained through them is good. In summary if the sensitivity or specificity or both are low, the efficiency of such classifier is also low and therefore such classifier is not desirable. However, when the sensitivity and specificity of a classifier are both high ($\geq .75$), it generates a good efficiency which is not lower than 75%. This is slightly more relaxed than the result in Jolayemi (2007) where sensitivity and specificity of 0.80 or higher was recommended. We observed that for low efficiency the variability measured by the standard deviation is higher compared to when efficiency is high. In general the standard deviation for the efficiency is insensitive to increasing sample sizes, that is not unexpected because the efficiency depicted similar invariant results with increasing sample sizes.

5. Numerical Example

The following data are taken from Pagano and Ganvreau (1993) and relate to the study to investigate the use of radionuclide Ventriculography in detecting coronary disease.

Table 5.1 : Results Of Screening Of 993 Subjects For Coronary Disease

COMPETING TEST	DISEASE STATUS		TOTAL
	Present	Absent	
Positive	302	80	382
Negative	179	372	551
TOTAL	481	452	993

The sensitivity of outcome of Radionuclide Ventriculography is

$$\hat{\theta}_1 = 302 / 481 = 0.628$$

The specificity of Radionuclide Ventriculography is

$$\hat{\theta}_2 = 372 / 452 = 0.823$$

However because sensitivity θ_1 was estimated as 0.628 (< 0.75) the use of radionuclide Ventriculography was not good enough even though the specificity was above 82%.

Thus the efficiency of this radionuclide Ventriculography was

$$\hat{\lambda} = (302 + 372) / 933 = 0.722$$

and not good enough to regard the use of radionuclide Ventriculography to detect coronary disease as acceptable.

5. Conclusion

For an efficient screening instrument, the competing classifier must have both high sensitivity and specificity (> 0.75) for such classifier to be desirable, otherwise we run the risk of high overall misclassification. The danger of misclassification are in two folds:

If a diseased subject is not identified by the test instrument, such a subject may suffer lack of early treatment(s) for a possible control or complete cure. Subsequently such subject(s) may die because of the error through the classifier. On the other hand, if a subject free of a disease is misclassified as having the disease (wrongly misclassified as HIV/AIDS positive for example) such individual may be exposed to an undesirable psychological trauma and/ or often times harmful therapy.

REFERENCES

1. Agresti, A. : "Categorical Data Analysis". Wiley, New York. (1990).
2. Birch, M. W. : "Maximum Likelihood In Three-Way Contingency Tables". J. Roy. Statist. Soc. B25, (1964), 220 - 233.
3. Jolayemi, E. T. : "Establishing An Equivalence/Noninferiority Test In Two Outcome Situation". Susan Conference, Kumasi, Ghana. (2007).
4. Chen, J. J.; Tsong, Y. and Kang, S. H. : "Tests For Equivalence On Non - Inferiority Between Two Proportions". Drug Information Journal, 34, (2000), 593 - 602.
5. Lawal, H. B. : "Categorical Data Analysis With Sas And Spss Application". Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey, London. (2003).
6. Mcneil, B. J.; Keeler, E. and Adelstein, S. J. : "Primer On Certain Elements Of Medical Decision Making". New England Journal Of Medicine. 239, (1975), 211- 215.
7. Nam, J. : "Sample Size Determination In Statistical Trials To Establish The Equivalence Of Two Treatments". Statistics In Medicine, 14, (1995), 2007- 2049.
8. Nam, J. : "Establishing Equivalence Of Two Treatments And Sample Size Requirements In Matched - Pairs Design". Biometrics, 53, (1997), 1422- 1430.
9. Sanni, O. O. M. : "A Study Of Some Goodness - Of -Fit Statistics For Contingency Tables With Sparse Data". Ph. D Thesis, University Of Ilorin. Unpublished. (1997).

10. Tang, N. S.; Tang, M. L. and Chan, I. . S. F. : “On Tests Of Equivalence Via Non-Unity Relative Risk For Matched-Pair Design”. *Statistics In Medicine*, In Press. (2002).
11. Yanagawa, T.; Tango, T. and Hiejima, Y. : “Mantel-Haenszel-Type Tests For Testing Equivalence Or More Than Equivalence In Comparative Clinical Trials”. *Biometrics*, 50, (1994), 659 - 864.