ROC analysis concepts

Christos T. Nakas

Laboratory of Biometry, University of Thessaly, 38446 Volos, Greece
Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland

Overview

1. Introduction
2. Diagnostic and Predictive Accuracy of Tests
3. ROC curve analysis
4. Some comments
5. Hands-on

Test results / measurements

The information used in the task of detection and prediction is obtained from the results of *tests*, which involve such modalities as:

- Devices (e.g. imaging technologies)
- Laboratory assays (e.g. laboratory measurements or molecular biomarkers)
- Algorithms (e.g. risk scores in medicine or prediction rules in machine learning)

Typology of test results

- A test result is said to be positive (T+) if it is indicative of disease. It is negative otherwise (T−).
- The test result may be contradictory to the true disease state.
- A diseased subject is indicated as D+, otherwise as D−.
- Test results can be
  Binary-scaled (e.g. T+, T−),
  Ordinal-scaled (e.g. Definitely non-diseased $[- \, -]$, Non-diseased $[-]$, Not Sure $[-/+]$, Diseased $[+]$, Definitely diseased $[+ \, +]$),
  Continuous-scaled (e.g. a measurement in $[0, 100]$).

Reference standards

- One needs to know in advance the true disease status of the subjects in the study.
- And -of course- the results of the diagnostic test under study.
- True disease status is termed as the `reference standard', used to reflect the fact that it is rare for the true disease state to be measured without any error whatsoever.

Accuracy and Classification

- *Accuracy*: the degree of success in detecting the true state of a target condition (*diagnostic accuracy*) or in predicting the true state of a target condition using available test information (*predictive accuracy*).
- The detection task is often called *classification* in settings beyond medicine.

Sensitivity and specificity - binary classification

- Evaluation of an antigen-based rapid test for the diagnosis of SARS-CoV-2 presence in respiratory samples.
- A total of 127 subjects were tested. The reference standard was RT-PCR, which resulted in 45 negative and 82 positive results.
- The test resulted in 50 negative and 77 positive results.

	Negative test: T$-$	Positive test: T$+$
D$-$	45	0
D$+$	5	77

- *Sensitivity* refers to 'positivity in disease', while *specificity* refers to 'negativity in health'.
- Estimated test sensitivity 93.9% (95% CI: 86.5% - 97.4%)
- Estimated test specificity equal to 100%.

Sensitivity and specificity in general

- For ordinal and continuous test results assume that higher values of the test variable indicate higher likelihood of the presence of the target condition,
- subsequently, for each threshold value c of T the test result is considered `positive' for the presence of the target condition if $T > c$ and `negative' otherwise.
- We can then define the classical twin measures of diagnostic test accuracy of *sensitivity* and *specificity* as the conditional probabilities $P(T > c|D = 1)$ and $P(T \leq c|D = 0)$ respectively.

Continuous-scaled marker

- Brucellosis and CD4 measurements
- CD4 measurements for the 15 Controls were:

$$\{59, 66, 45, 62, 51, 50, 49, 58, 53, 42, 50, 47, 51, 62, 48\}$$

- measurements for the 12 Acute Brucellosis Cases were:

$$\{72, 70, 69, 82, 68, 59, 76, 61, 59, 73, 49, 77\}.$$

- Suppose we choose $c = 61$, thus considering measurements larger than or equal to 61 as diseased and measurements smaller than 61 as healthy.
- Sensitivity is equal to 75% (i.e. 9/12)
- Specificity is 80% (i.e. 12/15).

PPV, NPV

Although the sensitivity and the specificity quantify the accuracy of the diagnostic marker under study, they do not answer to the major existential question of the subjects in the study:

- `Given that I get a positive test result, what is the probability that I really have the disease?',
- `Given that I get a negative test result, what is the probability that I really do not have the disease?'.
- These probabilities are defined as the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV) respectively.

PPV, NPV

- For a test positivity threshold c, the positive predictive value (PPV) is defined as the conditional probability $P(D = 1|T > c)$ and the negative predictive value (NPV) is defined as the conditional probability $P(D = 0|T \leq c)$.
- To calculate PPV, NPV one simply needs to apply Bayes' formula,

$$PPV(c) = \frac{sens(c) \cdot prob(D+)}{sens(c) \cdot prob(D+) + (1 - spec(c)) \cdot prob(D-)} \quad (1)$$

and

$$NPV(c) = \frac{spec(c) \cdot prob(D-)}{spec(c) \cdot prob(D-) + (1 - sens(c)) \cdot prob(D+)}, \quad (2)$$

where $prob(D+)$ is the prevalence of disease (i.e. proportion of diseased) *in the target population* and $prob(D-) = 1 - prob(D+)$.

Likelihood ratios

- The positive likelihood ratio (LR^+) is defined as $LR(c)^+ = \frac{sens(c)}{1-spec(c)}$.
- The negative likelihood ratio (LR^-) is defined as $LR(c)^- = \frac{1-sens(c)}{spec(c)}$.
- The latter quantities provide measures of the relative discriminatory performance of the diagnostic marker under study in the populations of diseased and non-diseased subjects.

Ordinal-scaled marker

- The reader assessed images from 98 subjects and gave a score from 1 (less indicative of disease) to 7 (more indicative of disease).
- The true status of each subject was evaluated with a biopsy.

True disease status *vs.* score	1	2	3	4	5	6	7
D−	50	3	11	5	1	0	1
D+	13	1	3	6	0	3	1

- cutoff $(>= 1)$, sens=100.00%, spec=0.00%
- cutoff $(>= 2)$, sens=51.85%, spec=70.42% ...
- cutoff (> 7), sens=0.00%, spec=100.00%

The ROC curve

- Clearly the sensitivity and specificity of the test depend on the positivity threshold c.
- A cut-off point, c, that will be used in practice in order to discriminate between non-diseased and diseased subjects needs to be defined that will serve for diagnosing disease based on the marker under consideration.
- As c varies over the range of possible values of T, the sensitivity and specificity of the test vary from 0 to 1.
- The pairs of *(sensitivity, 1-specificity)* obtained as c varies over the entire range of possible values of T constitute the ROC curve for T.

Understanding ROC curves

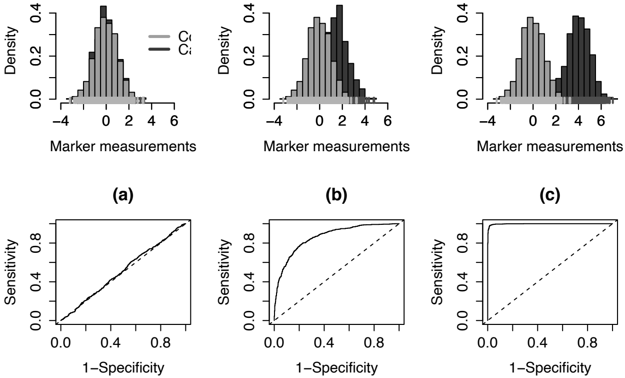
The Area Under the ROC Curve (AUC)

- The ROC curve displays the trade-off between test sensitivity and specificity across the range of values of the threshold for test positivity.
- A steep rise of the curve after the (0,0) point indicates that quick gains in sensitivity can be made for small loss in specificity.
- The Area Under the Curve (AUC) provides an overall measure of the attainable sensitivity values as specificity varies.
- The AUC reflects the amount of discrimination of the marker-level distributions since $AUC = P(X_1 < X_2)$.

Remarks

- Higher areas indicate better diagnostic accuracy for the test.
- The sensitivity and the specificity of the test are coupled quantities and need to be studied together even in situations in which practical interest is focused on one or the other.
- Generalizations of the ROC curve to settings in which the target condition takes more than two values have also been developed.

ROC curve demo

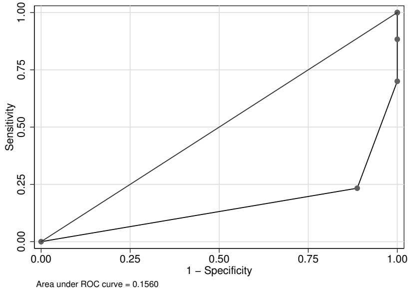


The maximum of the Youden index

- Another useful summary measure for the ROC curve is the maximum distance from the curve to the main diagonal of the unit square.
- This measure is equivalent to the maximum of the Youden index (defined as $J = max_c\{sens(c) + spec(c) - 1\}$).
- J can be used for the selection of an optimal operating cut-off point from a quantitative perspective, since it maximizes accuracy.

ROC general trivia

We have assumed that, in general, $X_2 > X_1$. In practice, the opposite direction may hold. In such a case, if one uses the definitions given until now, they will end up with an ROC curve lying under the main diagonal in the unit square. Also, AUC will be less than 0.5. Such a result simply flags that $X_1 > X_2$ and it is equally informative with the 'standard' case.



ROC general trivia

- LoD measurements can be readily accommodated for in ROC curve analysis. In the standard case, a mass of tied measurements will be considered that might be visible on the empirical ROC curve (or not).
- ROC curve analysis has been an active area of research in diverse scientific fields. Methodological developments and applications exist in virtually all scientific fields of applied research.

▸ Web-based Calculator for ROC Curves

Connections to logistic regression

- Although the ROC curve was introduced as a generalization of the notion of test sensitivity and specificity and is mostly used to assess *diagnostic accuracy*, the curve can also be used to assess the strength of prediction.
- For example, in a logistic regression model, the binary response can be considered as the "target condition" and the estimated probability of response as the test variable.
- The area of the resulting ROC curve is the value of the so-called *c-statistic* for the logistic regression model.

A note

Guidelines on correct reporting of diagnostic accuracy studies have been offered by the Standards for Reporting of Diagnostic Accuracy Studies (STARD) group and appear in the 'equator network' website: `http://www.equator-network.org`. These must be followed in order to ensure reliability and value of published ROC related health research literature.

The ROC curve for a continuous-scaled marker

- Suppose that we have measurements from 2 healthy and 3 diseased subjects, specifically, $X_{11} = 5, X_{12} = 7$ and $X_{21} = 6, X_{22} = 8, X_{23} = 9$.
- The cut-off point c can take on all possible values in the support of marker measurements (say 3 through 10) but ...
- only six different values for c are relevant, since any value of c greater than 9 will result in sensitivity equal to zero and specificity equal to one, any value of c greater than or equal to six and less than 7 will result in sensitivity equal to one and specificity equal to 0.5 etc.

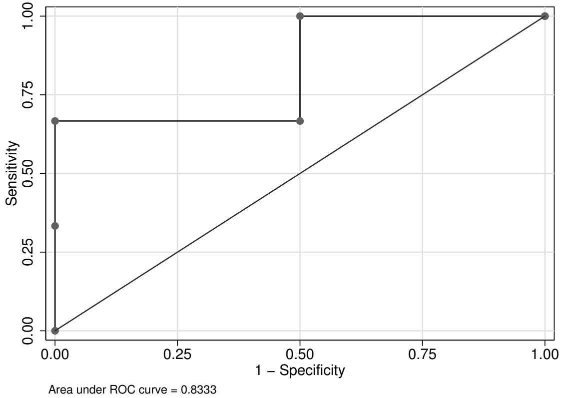
The ROC curve for a continuous-scaled marker

Cut-off points, sensitivity, specificity, and likelihood ratios for the simple example of 2 healthy vs. 3 diseased subjects.

Cut-off	Sensitivity	Specificity	LR^{+}	LR^{-}
(≥ 5)	1.000	0.000	1.000	
(≥ 6)	1.000	0.500	2.000	0.000
(≥ 7)	0.667	0.500	1.333	0.667
(≥ 8)	0.667	1.000		0.333
(≥ 9)	0.333	1.000		0.667
(> 9)	0.000	1.000		1.000

The ROC curve for a continuous-scaled marker

The ROC curve corresponding to the simple example with 2 healthy vs. 3 diseased subjects:



The ROC curve for an ordinal-scaled marker

- Measurements are characterizations {I: $--$, II: $-$, III: $-/+$, IV: $+$, V: $++$}.
- In practice six cut-off points can be defined (i.e. the number of categories+1) and the construction of the ROC curve is based on the six resulting (1-specificity, sensitivity) points.
- A specific characteristic of ROC curves for markers with ordinal data is the presence of ties between measurements of subjects from the two groups. This is not a problem of course, the only impact being diagonal segments instead of steps on the resulting graph of the ROC curve.

The ROC curve for an ordinal-scaled marker

Contingency table for a marker with ordinal test results by true disease state. Numbers represent frequencies.

	I:$- -$	II:$-$	III:$-/+$	IV:$+$	V:$+ +$
D$-$	2	4	1	1	0
D$+$	0	0	2	3	3

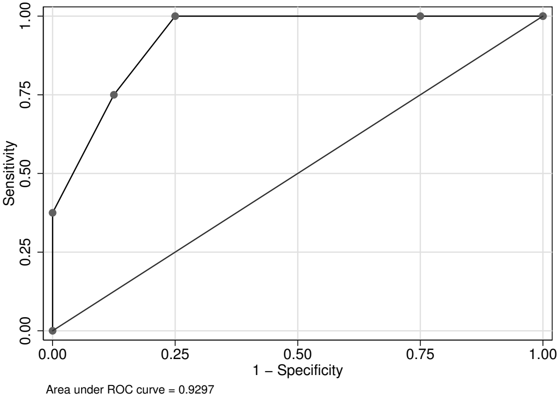
The ROC curve for an ordinal-scaled marker

Cut-off points, sensitivity, specificity, and likelihood ratios for the simple example of 8 healthy vs. 8 diseased subjects.

Cut-off	Sensitivity	Specificity	LR^+	LR^-
(≥ 1)	1.000	0.000	1.000	
(≥ 2)	1.000	0.250	1.333	0.000
(≥ 3)	1.000	0.750	4.000	0.000
(≥ 4)	0.750	0.875	6.000	0.286
(≥ 5)	0.375	1.000		0.625
(> 5)	0.000	1.000		1.000

The ROC curve for an ordinal-scaled marker

The ROC curve corresponding to the simple example with 8 healthy vs. 8 diseased subjects.



The Area Under the ROC curve (AUC)

- For the continuous-scaled marker simple example the $\tilde{AUC}$ is equal to 0.833 since it is the sum of two square areas equal to $(0.5 \times 0.667 + 0.5 \times 1.0)$.
- Similarly, for the ordinal-scaled marker simple example, $\tilde{AUC} = 0.93$.

References

- Krzanowski WJ, Hand D J (2009). *ROC curves for continuous data.* CRC Press.
- Pepe MS (2004). *The statistical evaluation of medical tests for classification and prediction.* Oxford University Press.
- Zhou XH, Obuchowski NA, McClish DK (2011). *Statistical Methods in Diagnostic Medicine, 2nd ed.* Wiley.
- Zou KH, Liu A, Bandos AI, Ohno-Machado L, Rockette H (2011). *Statistical evaluation of diagnostic performance: Topics in ROC analysis.* CRC Press.