ROC

ROC produces a receiver operating characteristic (ROC) curve.

Notation and Definitions

d_i	Actual state for case *i*, it is either positive or negative; positive usually means that a test detected some evidence for a condition to exist.
x_i	Test result score for case *i*.
n_{TP}	Number of true positive decisions
n_{FN}	Number of false negative decisions
n_{TN}	Number of true negative decisions
n_{FP}	Number of false positive decisions
Sensitivity	Probability of correctly identifying a positive
Specificity	Probability of correctly identifying a negative
C	Cutoff or criterion value on the test result variable
n_-	Number of cases with negative actual state
n_+	Number of cases with positive actual state
$n_{-=j}$	Number of true negative cases with test result equal to *j*.
$n_{+>j}$	Number of true positive cases with test result greater than *j*.
$n_{+=j}$	Number of true positive cases with test result equal to *j*.
$n_{-<j}$	Number of true negative cases with test result less than *j*.
Q_1	The probability that two randomly chosen positive state subjects will both get a more positive test result than a randomly chosen negative state subject.
Q_2	The probability that one randomly chosen positive state subject will get a more positive test result than two randomly chosen negative state subjects.

Construction of the ROC Curve

The ROC plot is merely the graph of points defined by sensitivity and (1 – specificity). Customarily, sensitivity takes the y axis and (1 – specificity) takes the x axis.

Computation of Sensitivity and Specificity

The ROC procedure fixes the set of cutoffs to be the set defined by the values half the distance between each successive pair of observed test scores, plus $max(x_i) + 1$ and $min(x_i) - 1$.

Given a set of cutoffs, the actual state values, and test result values, one can classify each observation into one of TP, FN, TN, and FP according to a classification rule. Then, the computation of sensitivity and specificity is immediate from their definitions.

Four classification or decision rules are possible:

(1) a test result is positive if the test result value is greater than or equal to C and that a test result is negative if the test result is less than C;

(2) a test result is positive if the test result value is greater than C and that a test result is negative if the test result is less than or equal to C;

(3) a test result is positive if the test result value is less than or equal to C and that a test result is negative if the test result is greater than C; and

(4) a test result is positive if the test result value is less than C and that a test result is negative if the test result is greater than or equal to C.

Specificity

Specificity is defined by

$$\frac{n_{\mathrm{TN}}}{n_{\mathrm{TN}} + n_{\mathrm{FP}}}$$

Sensitivity

Sensitivity is defined by

$$\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}} + n_{\mathrm{FN}}}$$

Interpolation of the Points

When the test result variable is a scale variable, the number of distinct test result values and thus the number of cutoff points tend to increase as the number of observations (or test results) increases. Theoretically, in the "limit" the pairs of sensitivity and (1 – specificity) values form a dense set of points in itself and in some continuous curve, the ROC curve. A continuous interpolation of the points may be reasonable in this sense.

Note: The domain of the test result variable need only be a positive-measure subset of the real line. For example, it could be defined only on (-1, 0] and (1, $+\infty$). As long as the variable is not discrete, the ROC curve will be continuous.

When the test result variable is an ordinal discrete variable, the points never become dense, even when there are countably infinite number of (ordinal discrete) values. Thus, a continuous interpolation may not be justifiable. But, when it is reasonable to assume there is some underlying or latent continuous variable, an interpolation such as a linear interpolation, though imprecise, may be attempted. From now on, the test result variable is assumed continuous or practically so.

The problem is related to having ties, but not the same. In the continuous case, when values are tied, they are identical but unique. In the ordinal case with the grouped/discretized continuous interpretation, values in some underlying continuous scale range may be grouped together and represented by a certain value, usually the mid range value. Those values are represented as if they were ties, but in fact they are a collection of unordered values. Now, even if each category/group contains only one observation, the problem still exists unless the observation's latent value is identical to the representing value of the observation.

Case 1: No ties between actual positive and actual negative groups

If there are ties within a group, the vertical/horizontal distance between the points is simply multiplied by the number of ties. If not, all the points are uniformly spaced within each of the vertical and horizontal directions, because as a cutoff value changes, only one observation at a time switches the test result.

Case 2: Some ties between actual positive and actual negative groups

For ties between actual positive and actual negative groups, both of the n_{TP} and n_{FP} change simultaneously, and we do not know "the correct path between two adjacent points" (Zweig and Campbell, 1993, p. 566). "It could be the minimal path (horizontal first, then vertical) or the maximal path (vice versa). The straight diagonal line segment is the average of the two most extreme paths and tends to underestimate the plot for diagnostically accurate test" (Zweig and Campbell, 1993, p. 566). But, it is our choice here. In passing, the distance and angle of this diagonal line depend on the numbers of ties within D+ and D- groups.

Computation of the Area Under the ROC Curve

The area under the ROC curve

Let x represent the scale of the test result variable, with its low values suggesting a negative result and the high values a positive result. Denote by x_+ the x values for cases with positive actual states. Similarly, denote by x_- the x values for cases with negative actual states. Then, the "true" area under the ROC curve is

$$\theta = \Pr(x_+ > x_-) .$$

The nonparametric approximation of θ is

$$W = \frac{1}{n_+ n_-} \sum_{\substack{\textit{all possible} \\ \textit{combinations} \\ \textit{of } (x_+, x_-)}} s(x_+, x_-),$$

where n_+ is the sample size of D+, n_- is the sample size of D-, and

$$s(x_+, x_-) = \begin{cases} 1 & \text{if } x_+ > x_- \\ 1/2 & \text{if } x_+ = x_- \\ 0 & \text{if } x_+ < x_- \end{cases} .$$

Note that W is the observed area under the ROC curve, which connects successive points by a straight line, i.e., by the trapezoidal rule.

An alternative way to compute W is as follows:

$$W = \frac{1}{n_+ n_-} \sum_{\substack{x \in \{\text{set of all test} \\ \text{result values}\}}} \left\{ n_{-=j} \times n_{+>j} + \frac{n_{-=j} \times n_{+=j}}{2} \right\}.$$

When a low value of X suggests a positive test result and a high value a negative test result

If a low value of X suggests a positive test result and a high a negative test result, compute W as above and then

$$W' = 1 - W,$$

where W' is the estimated area under the curve when a low test result score suggests a positive test result.

The SE of the area under the ROC curve statistic

Under the nonparametric (distribution-free) assumption

The standard deviation of W is estimated by:

$$\mathrm{SE}(W) = \sqrt{\frac{W(1-W) + (n_+ - 1)(\hat{Q}_1 - W^2) + (n_- - 1)(\hat{Q}_2 - W^2)}{n_+ n_-}}.$$

where

$$\hat{Q}_1 = \frac{1}{n_- n_+^2} \sum_x n_{-=j} \times [n_{+>j}^2 + n_{+>j} \times n_{+=j} + \frac{n_{+=j}^2}{3}]$$

and

$$\hat{Q}_2 = \frac{1}{n_-^2 n_+} \sum_x n_{+=j} \times [n_{-<j}^2 + n_{-<j} \times n_{-=j} + \frac{n_{-=j}^2}{3}].$$

When a low value of X suggests a positive test result and a high value a negative test result

If we assume that a low value of x suggests a positive test result and a high value a negative test result, then we estimate the standard deviation of W' by $\mathrm{SE}(W') = \mathrm{SE}(W)$.

Under the bi-negative exponential distribution assumption, given $n_+ = n_-$

Under the bi-negative exponential distribution assumption when $n_+ = n_-$,

$$\hat{Q}_1 = \frac{W}{2-W}$$

and

$$\hat{Q}_2 = \frac{2W^2}{1+W}.$$

$\mathrm{SE}(W)$ is then computed as before.

When a low value of x suggests a positive test result and a high value a negative test result

Once again, $\mathrm{SE}(W')$ = $\mathrm{SE}(W)$.

The asymptotic confidence interval of the area under the ROC curve

A 2-sided asymptotic $c\% = (100-\alpha)\%$ confidence interval for the true area under the ROC curve is

$$W \pm Z_\alpha \ \mathrm{SE}(W).$$

When a low value of x suggests a positive test result and a high value a negative test result

$$W' \pm Z_\alpha \ \mathrm{SE}(W')$$

The asymptotic *P*-value under the null hypothesis that $\theta = 0.5$ vs. the alternative hypothesis that $\theta \neq 0.5$

Since W is asymptotically normal under the null hypothesis that $\theta = 0.5$, we can calculate the asymptotic *P*-value under the null hypothesis that $\theta = 0.5$ vs. the alternative hypothesis that $\theta \neq 0.5$:

$$\Pr\left(|Z| > \left| \frac{W - 0.5}{\mathrm{SD}(W)\big|_{\theta=0.5}} \right| \right) = 2\Pr\left(Z > \left| \frac{W - 0.5}{\mathrm{SD}(W)\big|_{\theta=0.5}} \right| \right).$$

In the nonparametric case,

$$\mathrm{SD}(W)\big|_{\theta=0.5} = \sqrt{\frac{\theta(1-\theta) + (n_+ - 1)(Q_1 - \theta^2) + (n_- - 1)(Q_2 - \theta^2)}{n_+ n_-}}\Bigg|_{\theta=0.5}$$

$$= \sqrt{\frac{0.5(1-0.5) + (n_+ - 1)(1/3 - 0.5^2) + (n_- - 1)(1/3 - 0.5^2)}{n_+ n_-}}$$

$$\left(= \sqrt{\frac{n_+ + n_- + 1}{12 n_+ n_-}} = \sqrt{\frac{n_+ n_- (n_+ + n_- + 1)}{12}} \Big/ n_+ n_- \right),$$

because we can deduce that $Q_1 = 1/3$ and $Q_2 = 1/3$ under the null hypothesis that $\theta = 0.5$. The argument for $Q_1 = 1/3$ is as follows. $\theta = 0.5$ implies that the distribution of test results of positive actual state subjects is identical to the distribution of test results of negative actual state subjects. So, the mixture of the two distributions is identical to either one of the distributions. Then, we can reinterpret Q_1 as the probability that, given three randomly chosen subjects from the (mixture) distribution, the subject with the lowest test result was selected, say, first. (One may consider this subject as a negative state subject and the other two as positive state subjects.) From here on, we can pursue a purely combinatorial argument, irrespective of the distribution of subjects' test results, because the drawings are independent and given. There are $3! = 3 \times 2 \times 1 = 6$ ways to order the three subjects, and there are two ways in which the subject with the lowest test result comes first. So, if $\theta = 0.5$, $Q_1 = 2/6 = 1/3$. The argument for $Q_2 = 1/3$ is similar.

In the bi-negative exponential case,

$$\begin{aligned}
\mathrm{SD}(W)\big|_{\theta=0.5} &= \sqrt{\frac{\theta(1-\theta)+(n_+-1)(Q_1-\theta^2)+(n_--1)(Q_2-\theta^2)}{n_+n_-}}\Bigg|_{\theta=0.5} \\
&= \sqrt{\frac{\theta(1-\theta)+(n_+-1)(\frac{\theta}{2-\theta}-\theta^2)+(n_--1)(\frac{2\theta^2}{1+\theta}-\theta^2)}{n_+n_-}}\Bigg|_{\theta=0.5} \\
&= \sqrt{\frac{0.5(1-0.5)+(n_+-1)(\frac{0.5}{2-0.5}-0.5^2)+(n_--1)(\frac{2\times 0.5^2}{1+0.5}-0.5^2)}{n_+n_-}} \\
&= \sqrt{\frac{0.5(1-0.5)+(n_+-1)(1/3-0.5^2)+(n_--1)(1/3-0.5^2)}{n_+n_-}},
\end{aligned}$$

where $n_+ = n_-$. (Note that this formula is identical to the nonparametric one except for the sample size restriction.)

When a low value of x suggests a positive test result and a high value a negative test result

The asymptotic *P*-value under the null hypothesis that $\theta' = 0.5$ vs. the alternative hypothesis that $\theta' \neq 0.5$, if desired, may be computed, using W' and $\mathrm{SD}(W')\big|_{\theta=0.5} = \mathrm{SD}(W)\big|_{\theta=0.5}$.

References

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology*, **12**, 387-415.

Beck, J. Robert and Shultz, Edward (1986). The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch. Pathol. Lab. Med.*, **110**, 13-20.

Centor, R. M. and Schwartz, J. S. (1985). An evaluation of methods for estimating the area under the receiver operating characteristic (ROC) curve. *Med. Decis. Making*, **5**, 149-156.

Dorfman, Donald D. and Alf, Edward, Jr. (1968). Maximum likelihood estimation of parameters of signal detection theory—A direct solution. *Psychometrika*, **33**, 117-124.

Dorfman, Donald D. and Alf, Edward, Jr. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—Rating-method data. *Journal of Mathematical Psychology*, **6**, 487-496.

Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. John Wiley and Sons, New York.

Griner, Paul F., Mayewski, Raymond J., Mushlin, Alvin I., and Greenland, Philip (1981). Selection and interpretation of diagnostic tests and procedures: Principles and applications. *Annals of Internal Medicine*, **94**, 553-600.

Hanley, James A. and McNeil, Barbara J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.

Hanley, James A. and McNeil, Barbara J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, **148**, 839-843.

Metz, Charles E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8**, 283-298.

Metz, Charles E. (1993). *ROC Software Package.*

Metz, Charles E. and Kronman, Helen B. (1980). Statistical significance tests for binormal ROC curves. *Journal of Mathematical Psychology*, **22**, 218-243.

Schoonjans, Frank (1993-1998). *MedCalc Version 4.20.011*.

Woods, Kevin and Bowyer, Kevin W. (1997). Generating ROC curves for artificial neural networks. *IEEE Transactions on Medical Imaging*, **16**, 329-337.

Zweig, Mark H. and Campbell, Gregory (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561-577.