Receiver Operating Characteristic Curves

Elizabeth Hill, PhD
Associate Professor of Biostatistics
Hollings Cancer Center
Medical University of South Carolina
hille@musc.edu

HCC Cancer Control Program Journal Club
08 January 2015

Outline

- Adapted primarily from *The Statistical Evaluation of Medical Tests for Classification and Prediction* (2003). MS Pepe, Oxford University Press, NY.
- Prediction and classification tests
- Assessing performance of a binary test
- Assessing performance of a continuous test
- Discussion of AJE article

Prediction and classification

- Diagnosis: disease versus non-disease
- Screening: for early diagnosis
- Prognosis: predicting outcome
- Treatment selection: predict treatment response
- Tests can be based on
    - clinical signs or symptoms
    - laboratory tests
    - imaging technology
    - genomic/proteomic/metabolomic signature
    - antibody arrays

Measuring classifier performance of a binary test

Definitions

Binary Test: $Y = 1$ (positive) 0 (negative)
Disease: $D = 1$ (present) 0 (absent)

Classification of test results by disease status

	$D = 0$	$D = 1$
$Y = 0$	True negative	False negative
$Y = 1$	False positive	True positive

Classification probabilities

True positive fraction = TPF = $P[Y = 1|D = 1]$ = Sensitivity
False positive fraction = FPF = $P[Y = 1|D = 0]$ = 1 - Specificity

Measuring classifier performance of a continuous test

- Assume larger values of $Y \Rightarrow D = 1$
- Thresholding positivity rule - "$Y \geq c$"
- TPF(c) = $P[Y \geq c | D = 1]$ = Sensitivity(c)
  FPF(c) = $P[Y \geq c | D = 0]$ = 1 - Specificity(c)
- Receiver Operating Characteristic (ROC) Curve

ROC definition

$$\text{ROC}(c) = \{(\text{FPF}(c), \text{TPF}(c)), c \in \mathcal{R}\}$$

Historical context

"ROC analysis is part of a field called 'Signal Detection Theory' developed during World War II for the analysis of radar images. Radar operators had to decide whether a blip on the screen represented an enemy target, a friendly ship, or just noise. Signal detection theory measures the ability of radar receiver operators to make these important distinctions. Their ability to do so was called the Receiver Operating Characteristics. It was not until the 1970s that signal detection theory was recognized as useful for interpreting medical test results." (*Practical Graph Mining with R*, p.391. CRC Press, 2014)

Simple logistic regression example

- $\pi = P[D = 1]$
- $\text{logit}(\pi) = \beta_0 + \beta_1 Y$, Y continuous
- $\pi = \frac{1}{1+e^{-(\beta_0+\beta_1 Y)}}$ (inverse logit)
- Large $\pi \Rightarrow D = 1$ so we can use π instead of Y for classification

Subject	D	$\hat{\pi}$
1	1	0.95
2	1	0.95
3	0	0.72
4	1	0.72
5	1	0.72
6	0	0.24
7	1	0.12
8	0	0.08
9	0	0.08
10	0	0.04

$c = 1.0$ (or any c larger than 0.95)

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 1.0$?
1	1	0.95	No
2	1	0.95	No
3	0	0.72	No
4	1	0.72	No
5	1	0.72	No
6	0	0.24	No
7	1	0.12	No
8	0	0.08	No
9	0	0.08	No
10	0	0.04	No

- $\text{TPF}(1.0) = P[\hat{\pi} \geq 1.0 | D = 1] = 0/5 = 0.0$
- $\text{FPF}(1.0) = P[\hat{\pi} \geq 1.0 | D = 0] = 0/5 = 0.0$

$c = 0.95$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.95$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	No
4	1	0.72	No
5	1	0.72	No
6	0	0.24	No
7	1	0.12	No
8	0	0.08	No
9	0	0.08	No
10	0	0.04	No

- $\text{TPF}(0.95) = P[\hat{\pi} \geq 0.95 | D = 1] = 2/5 = 0.4$
- $\text{FPF}(0.95) = P[\hat{\pi} \geq 0.95 | D = 0] = 0/5 = 0.0$

$c = 0.72$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.72$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	Yes
4	1	0.72	Yes
5	1	0.72	Yes
6	0	0.24	No
7	1	0.12	No
8	0	0.08	No
9	0	0.08	No
10	0	0.04	No

- $\text{TPF}(0.72) = P[\hat{\pi} \geq 0.72 | D = 1] = 4/5 = 0.8$
- $\text{FPF}(0.72) = P[\hat{\pi} \geq 0.72 | D = 0] = 1/5 = 0.2$

$c = 0.24$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.24$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	Yes
4	1	0.72	Yes
5	1	0.72	Yes
6	0	0.24	Yes
7	1	0.12	No
8	0	0.08	No
9	0	0.08	No
10	0	0.04	No

- $\text{TPF}(0.24) = P[\hat{\pi} \geq 0.24 | D = 1] = 4/5 = 0.8$
- $\text{FPF}(0.24) = P[\hat{\pi} \geq 0.24 | D = 0] = 2/5 = 0.4$

$c = 0.12$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.12$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	Yes
4	1	0.72	Yes
5	1	0.72	Yes
6	0	0.24	Yes
7	1	0.12	Yes
8	0	0.08	No
9	0	0.08	No
10	0	0.04	No

- $\text{TPF}(0.12) = P[\hat{\pi} \geq 0.12 | D = 1] = 5/5 = 1.0$
- $\text{FPF}(0.12) = P[\hat{\pi} \geq 0.12 | D = 0] = 2/5 = 0.4$

$c = 0.08$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.08$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	Yes
4	1	0.72	Yes
5	1	0.72	Yes
6	0	0.24	Yes
7	1	0.12	Yes
8	0	0.08	Yes
9	0	0.08	Yes
10	0	0.04	No

- $\text{TPF}(0.08) = P[\hat{\pi} \geq 0.08 | D = 1] = 5/5 = 1.0$
- $\text{FPF}(0.08) = P[\hat{\pi} \geq 0.08 | D = 0] = 4/5 = 0.8$

$c = 0.04$

Subject	D	$\hat{\pi}$	$\hat{\pi} \geq 0.04$?
1	1	0.95	Yes
2	1	0.95	Yes
3	0	0.72	Yes
4	1	0.72	Yes
5	1	0.72	Yes
6	0	0.24	Yes
7	1	0.12	Yes
8	0	0.08	Yes
9	0	0.08	Yes
10	0	0.04	Yes

- $\text{TPF}(0.04) = P[\hat{\pi} \geq 0.04 | D = 1] = 5/5 = 1.0$
- $\text{FPF}(0.04) = P[\hat{\pi} \geq 0.04 | D = 0] = 4/5 = 1.0$

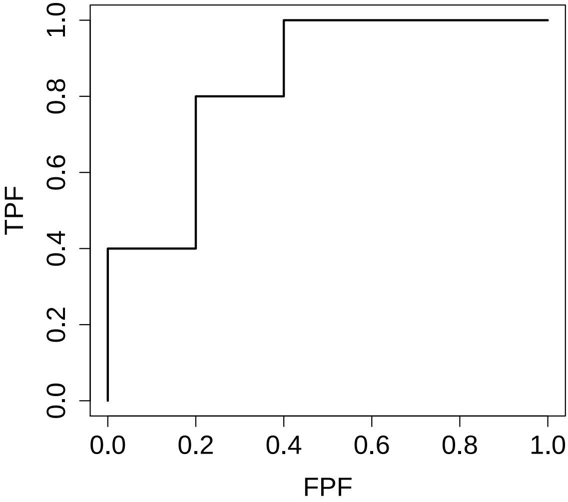
Estimated ROC curve

ROC definition

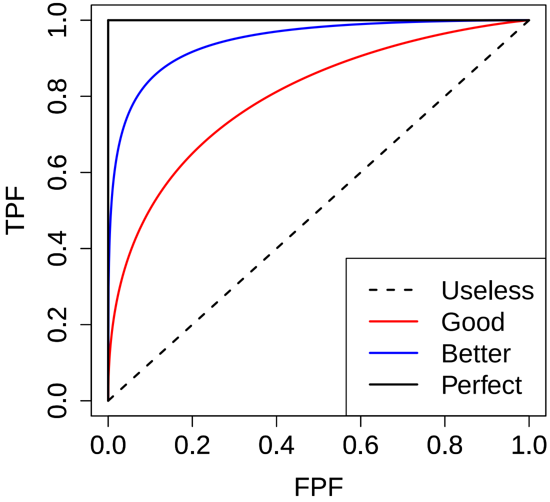
$$\text{ROC}(c) = \{(\text{FPF}(c), \text{TPF}(c)), c \in \mathcal{R}\}$$

c	FPF(c)	TPF(c)
1.0	0.0	0.0
0.95	0.0	0.4
0.72	0.2	0.8
0.24	0.4	0.8
0.12	0.4	1.0
0.08	0.8	1.0
0.04	1.0	1.0

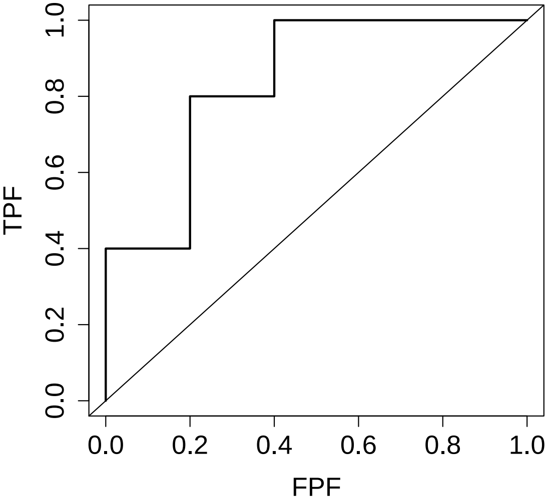
Estimated ROC curve (cont.)



Theoretical ROC curves



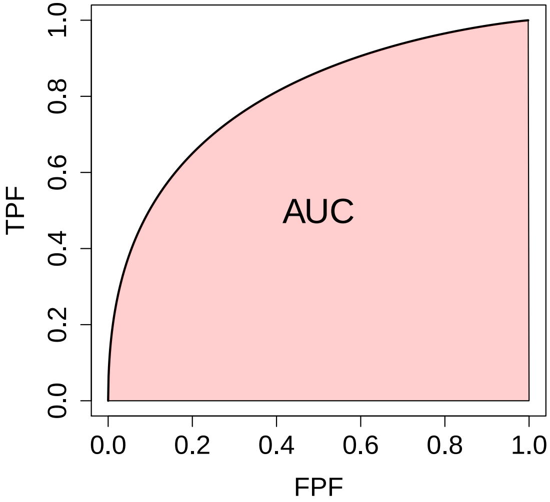
Estimated ROC curve with reference line



Attributes of ROC curve

- Provides complete description of potential performance
- Facilitates comparing and combining information across studies of the same test
- Facilitates comparing different tests on a common relevant scale
- Guides the choice of threshold in applications

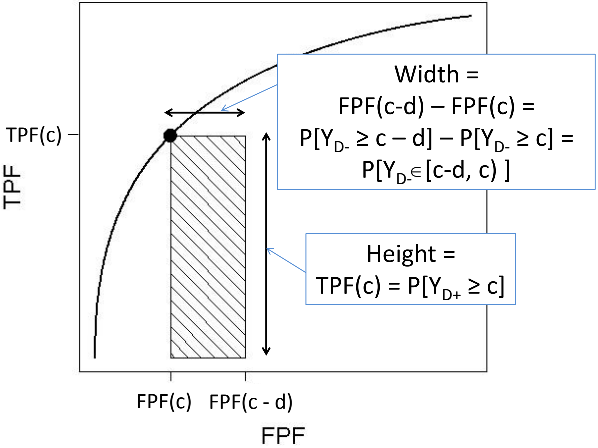
Area Under the Curve



AUC summary

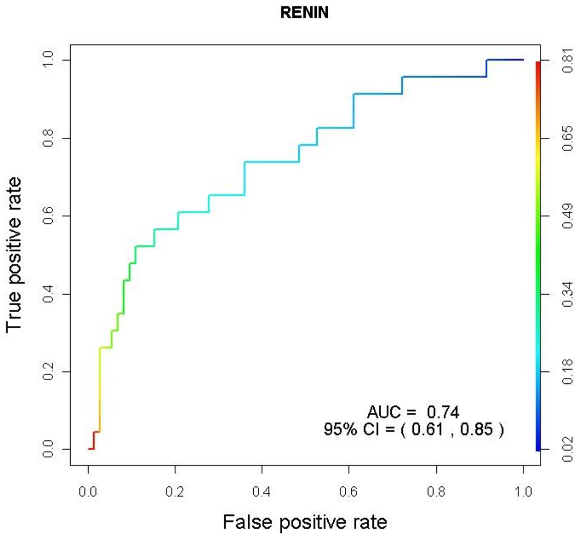
- AUC = $\int_0^1 \text{ROC}(t)dt$
- AUC = 0.5 $\Rightarrow$ useless test
- AUC = 1 $\Rightarrow$ perfect test
- AUC = Probability that test correctly 'orders' randomly selected diseased and non-disease subjects
- AUC $= P[(Y|D=1) > (Y|D=0)] = P[Y_{D+} > Y_{D-}]$

AUC interpretation - informal proof



- Area of rectangle = $P[Y_{D+} \geq c] \cdot P[Y_{D-} \in [c-d, c)]$ by independence of Y_{D+} and Y_{D-}
- Summation across all rectangles yields $P[Y_{D+} > Y_{D-}]$

Example using R library `ROCR`



References and resources

- R library `ROCR` - nice graphics
- R library `pROC` - compares ROC curves and partial ROC curves
- Bioconductor library `ROC`
- *The Statistical Evaluation of Medical Test for Classification and Prediction* by MS Pepe (2003). Oxford University Press.
- *Statistical Methods in Diagnostic Medicine, 2nd Edition* by X Zhou, NA Obuchowski and DK McClish (2011). Wiley.
- *Clinical Prediction Models* by EW Steyerberg (2010). Springer.