Probability-Class 3

January 14, 2014 Debdeep Pati

Predictive values

Predictive value positive (PV+): P(disease | test+)
Predictive value negative (PV-): P(no disease | test-)

Example. A: mammogram positive, B: developing breast cancer in next 2 years
Suppose that 7% of the general population of women will have a positive mammogram. What is the probability of developing breast cancer over the next 2 years among women in the general population?
P(breast cancer | mammogram+) = .1
P(breast cancer | mammogram-) = .0002
P(B) = P(breast cancer)
= P(breast cancer | mammogram+)P(mammogram+) + P(breast cancer | mammogram-)P(mammogram-) = .1(.07) + .0002(.93) = 0.00719
PV+ = P(breast cancer|mammogram+) = .1
PV- = P(no breast cancer | mammogram-) = 1- P(breast cancer | mammogram-) = 1-.0002 = .9998

Sensitivity and specificity

Sensitivity of a symptom is the probability that the symptom is present given that the person has a disease = P(symptom | disease)
Specificity of a symptom is the probability that the symptom is not present given that the person does not have a disease = P(no symptom | no disease)
A false negative is defined as a person who tests out as negative but who is actually positive.
A false positive is defined as a person who tests out at positive but who is actually negative.

Bayes' Rule

Let A = symptom and B = disease. Then

$$PV+ = P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

This can be written as

$$PV+ = \frac{sensitivity \times x}{sensitivity \times x + (1 - specificity) \times (1 - x)}$$

where $x = P(B)$ = probability of disease in the reference population.
Example: (Cancer) Suppose the disease is lung cancer and the symptom is cigarette smoking. If we assume 90% of people with lung cancer and 30 % of people without lung cancer are smokers, What is the sensitivity and specificity? Symptom: smoking, Disease: lung cancer
Sensitivity = P(symptom | disease) = .9
Specificity = P(no symptom | no disease) = 1- P(symptom | no disease) = .7

Example: (Hypertension) Suppose 84% of hypertensive and 23% of normotensives are classified as hypertensive by an automated blood-pressure machine. What are the predictive value positive and predictive value negative of the machine, assuming 20% of the adult population is hypertensive? The sensitivity = P(symptom | disease) = .84 and specificity = P(no symptom | no disease) = 1-.23 = .77. From Bayes rule PV+ = (sensitivity × x)/(sensitivity× x + (1-specificity)× (1-x))
PV- = (specificity × (1-x))/(specificity ×(1-x)+ (1-sensitivity) × x)
PV+ = (.84)(.2)/[(.84)(.2) + (.23)(.8)] = .168/.352 = .48
PV- = (.77)(.8)/[(.77)(.8) + (.16)(.2)] = .616/.648 = .95

An example in radiology

CT rating by radiologist

True Disease status	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	Total
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

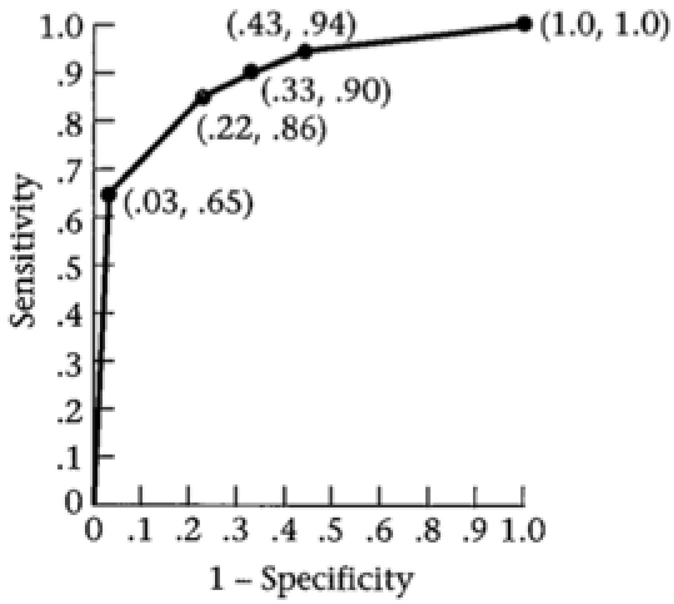
Table 1: Sensitivity vs. Specificity for different test criteria

Test positive criteria	Sensitivity	Specificity
1+	1	0
2+	0.94	0.57
3+	0.90	0.67
4+	0.86	0.78
5+	0.65	0.97
6+	0	1.0

Receiving operating characteristic (ROC) curve

ROC curve is a plot of the sensitivity versus (1-specificity) of a screening test, where the different points on the curve correspond to different cutoff points used to designate test positive.

Figure 3.7 **ROC curve for the data in Table 3.4***



*Each point represents (1 – specificity, sensitivity) for different test-positive criteria.

Practical

http://www.amstat.org/publications/jse/v13n2/datasets.kahn.html