Bayesian Reasoning

6.873/HST951
Peter Szolovits

Simplest Example

- Relationship between a diagnostic conclusion and a diagnostic test

	Disease Present	*Disease Absent*	
Test Positive	True Positive	False Positive	TP+FP
Test Negative	False Negative	True Negative	FN+TN
	TP+FN	FP+TN	

Definitions

	Disease Present	*Disease Absent*	
Test Positive	True Positive	False Positive	TP+FP
Test Negative	False Negative	True Negative	FN+TN
	TP+FN	FP+TN	

Sensitivity (true positive rate): TP/(TP+FN)

False negative rate: 1-Sensitivity = FN/(TP+FN)

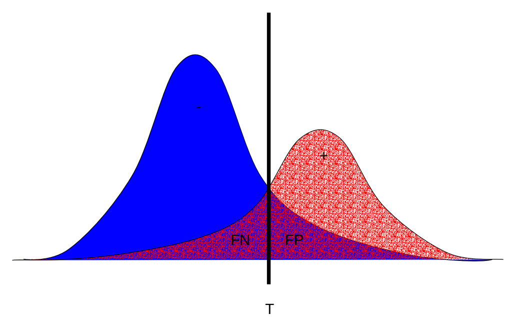
Specificity (true negative rate): TN/(FP+TN)

False positive rate: 1-Specificity = FP/(FP+TN)

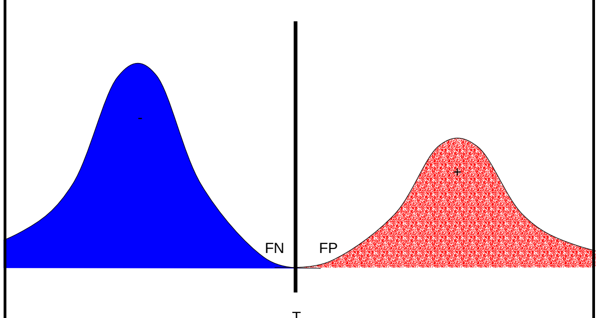
Positive Predictive Value: TP/(TP+FP)

Negative Predictive Value: TN/(FN+TN)

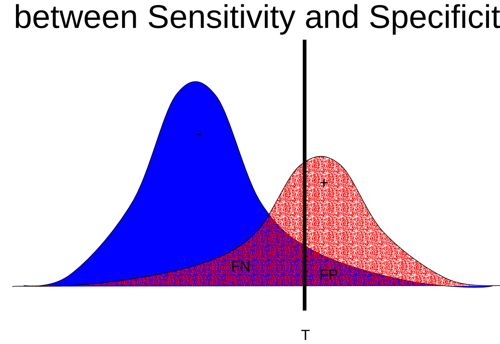
Test Thresholds



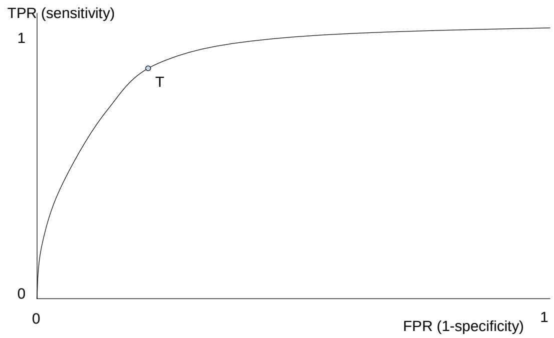
Wonderful Test



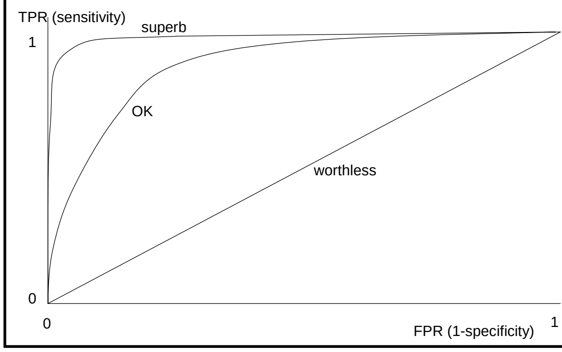
Test Thresholds Change Trade-off between Sensitivity and Specificity



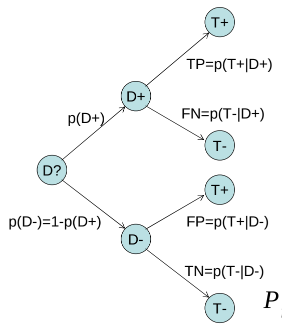
Receiver Operator Characteristic (ROC) Curve



What makes a better test?



How certain are we after a test?



Bayes' Rule:

$$P_{i+1}(D_j) = \frac{P_i(D_j)P(S|D_j)}{\sum_{k=1}^{n} P_i(D_k)P(S|D_k)}$$

Rationality

- Behavior is a continued sequence of choices, interspersed by the world's responses
- Best action is to make the choice with the greatest *expected value*
- … decision analysis

Example: Acute Renal Failure

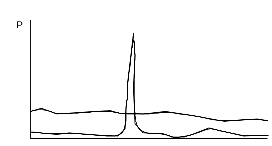
- Based on Gorry, *et al.*, AJM 55, 473-484, 1973.
- Choice of a handful (8) of therapies (antibiotics, steroids, surgery, etc.)
- Choice of a handful (3) of invasive tests (biopsies, IVP, etc.)
- Choice of 27 diagnostic "questions" (patient characteristics, history, lab values, etc.)
- Underlying cause is one of 14 diseases
  - We assume one and only one disease

Entropy of a distribution

$$H_i(P_1,\ldots,P_n) = \sum_{j=1}^{n} -P_j \log_2 P_j$$

For example:
H(.5, .5) = 1.0
H(.1, .9) = 0.47
H(.01, .99) = 0.08
H(.001, .999) = 0.01

H(.33, .33, .33) = 1.58 (!)
H(.005, .455, .5) = 1.04
H(.005, .995, 0) = 0.045



(!) -- should use $\log_n$

Interacting with ARF in 1973

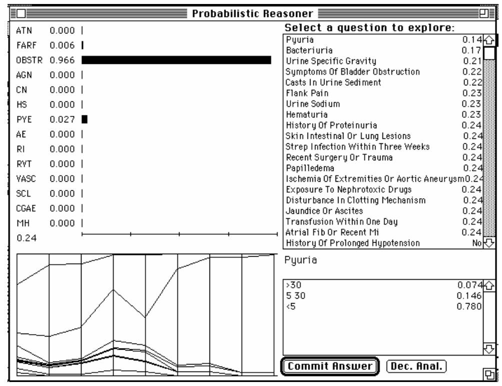
Question 1: What is the patient's age?
1 0-10
2 11-30
3 31-50
4 51-70
5 Over 70
Reply: 5

The current distribution is:

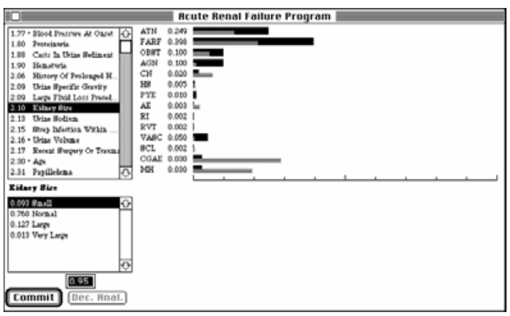
Disease	Probability
FARF	0.58
IBSTR	0.22
ATN	0.09

Question 2: What is the patient's sex?
1 Male
2 Pregnant Female
3 Non-pregnant Female
Reply: 1
. . .

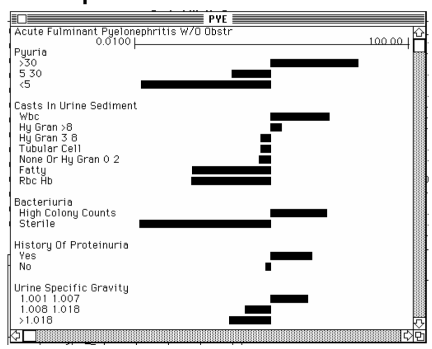
ARF in 1994



Local Sensitivity Analysis



Case-specific Likelihood Ratios



Assumptions in ARF

- Exhaustive, mutually exclusive set of diseases
- Conditional independence of all questions, tests, and treatments
- Cumulative (additive) disutilities of tests and treatments
- Questions have no modeled disutility, but we choose to minimize the number asked anyway

DeDombal, *et al.* Experience 1970's & 80's

- "Idiot Bayes" for appendicitis
- 1. Based on expert estimates -- *lousy*
- 2. Statistics -- *better than docs*
- 3. Different hospital -- *lousy again*
- 4. Retrained on local statistics -- *good*

Probabilistic Models

- What to represent?
  - Disease
  - Finding (signs, symptoms, labs, radiology, …)
  - Syndromes
  - History, predisposing conditions
  - Treatments
    - modify disease, cause new symptoms, …
  - (Outcomes, preferences, …)

State Space

- Set of random variables
- Possible values of each
- Assignment of probability to every possible combination of values of all variables
- p(v1=a1, v2=a2, v3=a3, …)

Questions of Interest

- Given a set of values of certain variables, what is the probability that certain other variables have certain other values?
- E.g., p(v1=a1, v7=a7|v2=a2, v4=a4)
  =p(v1=a1, v7=a7, v2=a2, v4=a4)
  /p(v2=a2, v4=a4)
- We *don't care* about all other variables
  - marginalize; i.e., sum over them all

Computational Cost

- For *n* binary variables, we need probability assignments to 2^n states.
- In programs such as DXPLAIN, *n* is on the order of thousands.
- Need to be *very* careful and clever
  - simple models
  - approximate solution techniques

Independence

- Two random variables are *independent* iff p(A&B)=p(A)p(B)
- Usually, however, variables may depend on others, but we are still interested whether they have a conditional dependence
- Two random variables are *conditionally independent* if for a conditioning variable D, p(A&B|D)=p(A|D)p(B|D)

Independence was crucial to ARF

- Diseases were *dependent*; mutually exhaustive and exclusive.
- Questions were conditionally independent, given disease.

ARF model convenient

- Odds: O(D)=P(D)/P(~D)=P(D)/(1-P(D))
- Likelihood ratio: L(S|D)=P(S|D)/P(S|~D)
- Bayes:O(D|S)=O(D)L(S|D)
- Multiple evidence:
- O(D|S1&S2&…)=O(D)L(S1|D)L(S2|D)…
- Log transform:
- W(D|S1&S2&…) =W(D)+W(S1|D)+W(S2|D)+…

Side comment on likelihood ratio

- L(s|d)=p(s|d)/p(s|~d) is constant only if ~d is a "fixed" entity
- If, as in ARF, we have d1, d2, d3, …, then p(s|~d_j)= $\sum_{i\neq j} p(d_i)\,p(s\,|\,d_i)$
- As probabilities vary over the d_j, p(s|~ d_j) will vary!

What if we made no assumptions in ARF?

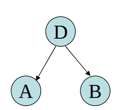
- Any combination of diseases: 2^14=16K
- Distinct probability for any combination of answers to any questions: p(q1=a12&q2=a24&…) 3^27=7.7*10^12
- p(q1=a12&q2=a24&…|d1&d2&~d3&…) 2^14*3^27 = 1.25*10^17, just for ARF
- Simplification is essential!

Conditional Independence is not Independence

- P(b&c|a)=P(a&b&c)/P(a)
- But P(b&c)=p(a)p(b&c|a)+p(~a)p(b&c|~a)

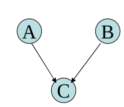
Conditional independence is not independence

- Information may still "flow" from one observation to another, even if they are conditionally independent given a disease, unless the disease is known with certainty
- p(D)=.2
- p(A|D)=.8, p(A|~D)=.1
- p(B|D)=.6, p(B|~D)=.1
- *a priori,* p(A)=.16+.08=.24, p(B)=.12+.04=.16
- P(D|A)=.67
- P(B|A)=.40+.03=.43, not .16!
- But, if p(D)=0 or 1, no effect.

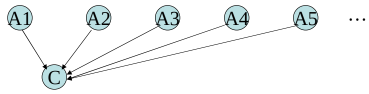


Conditional independence is not independence

- Two variables may be made *conditionally dependent* when we learn about a common descendant
- p(A)=.2, p(B)=.1
- p(C|A&B)=.8, p(C|A&~B)=.4
- p(C|~A&B)=.6, p(C|~A&~B)=.1
- p(C)=.2*.1*.8+.2*.9*.4+.8*.1*.6+.8*.9*.1=.208
- If we observe C, p(A&B)=.02*.8/.208=.077, but p(A)=.42, p(B)=.31. p(a&b|c) neq p(a|c)*p(b|c)



We don’t want to model high-arity dependence

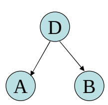


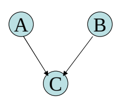
- P(C|A1&~A2&~A3&A4&…)
- too many probabilities
- Can we simplify?
  - Noisy or
  - noisy and
  - noisy max/min
  - ?

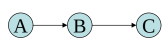
Noisy or

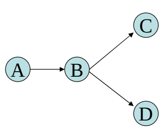
- p(C|A&B)=p(“C happened because of A”)
  * p(“C happened because of B”)
  * p(“C happened anyway”)
- (1-p(C|A&B)) =
  (1-p_c(C|A))*(1-p_c(C|B))*(1-*L*)
- p_c(C|A)<=p(C|A)

Simple Models
(Singly-Connected)





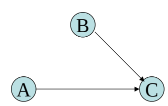


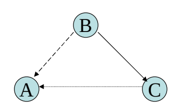


How to propagate likelihood information?
Like likelihood ratios.
(Pearl poly-tree algorithm)

Causality?

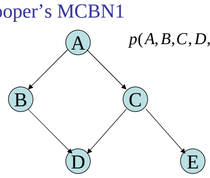
- Noisy-or (and, …)
- Bayes arrows are/are not causal
  - reversing an arrow adds new dependencies





General Case Bayes Nets
(Multiply-connected DAGs)

Cooper’s MCBN1



$$p(A,B,C,D,E) = p(A)\,p(B\mid A)\,p(C\mid A)\,p(D\mid B,C)\,p(E\mid C)$$

$$p(A=t,B=f,C=t,D=t,E=f)$$
$$= p(A=t)\,p(B=f\mid A=t)\,p(C=t\mid A=t)$$
$$p(D=t\mid B=f,C=t)\,p(E=f\mid C=t)$$

But what is $p(E=t\mid B=t)$?

$p(E{=}t|B{=}t)=p(E{=}t,B{=}t)/\,p(B{=}t)$

Consider:

$$p(E=t,B=t) = \sum_{A,C,D} p(A,B=t,C,D,E=t)$$

$$= \sum_{A,C,D} P(A)P(B=t\mid A)\,p(C\mid A)\,p(D\mid B=t,C)\,p(E=t\mid C)$$

Factoring to Simplify Computation

$$p(E=t,B=t) = \sum_{A,C,D} p(A,B=t,C,D,E=t)$$
$$= \sum_{A,C,D} P(A)P(B=t\mid A)\,p(C\mid A)\,p(D\mid B=t,C)\,p(E=t\mid C)$$
$$= \sum_{C} p(E=t\mid C)\left(\sum_{A} p(A)\,p(C\mid A)\,p(B=t\mid A)\right)\left(\sum_{D} p(D\mid B=t,C)\right)$$

In this simple example, 12 instead of 32 multiplications

How to build Bayes Nets?

- David Heckerman, *Pathfinder/ Intellipath*, around 1990