Simple Probabilistic Reasoning

6.873/HST951

Harvard-MIT Division of Health Sciences and Technology
HST.951J: Medical Decision Support

Change over 30 years

- 1970's: human knowledge, not much data
- 2000's: vast amounts of data, traditional human knowledge (somewhat) in doubt

- Could we "re-discover" all of medicine from data? *I think not!*
- Should we focus on methods for reasoning with uncertain data? *Absolutely!*
- But: Feinstein, A. R. (1977). "Clinical Biostatistics XXXIX. The Haze of Bayes, the Aerial Palaces of Decision Analysis, and the Computerized Ouija Board." <u>Clinical Pharmacology and Therapeutics</u> **21**: 482-496.

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)

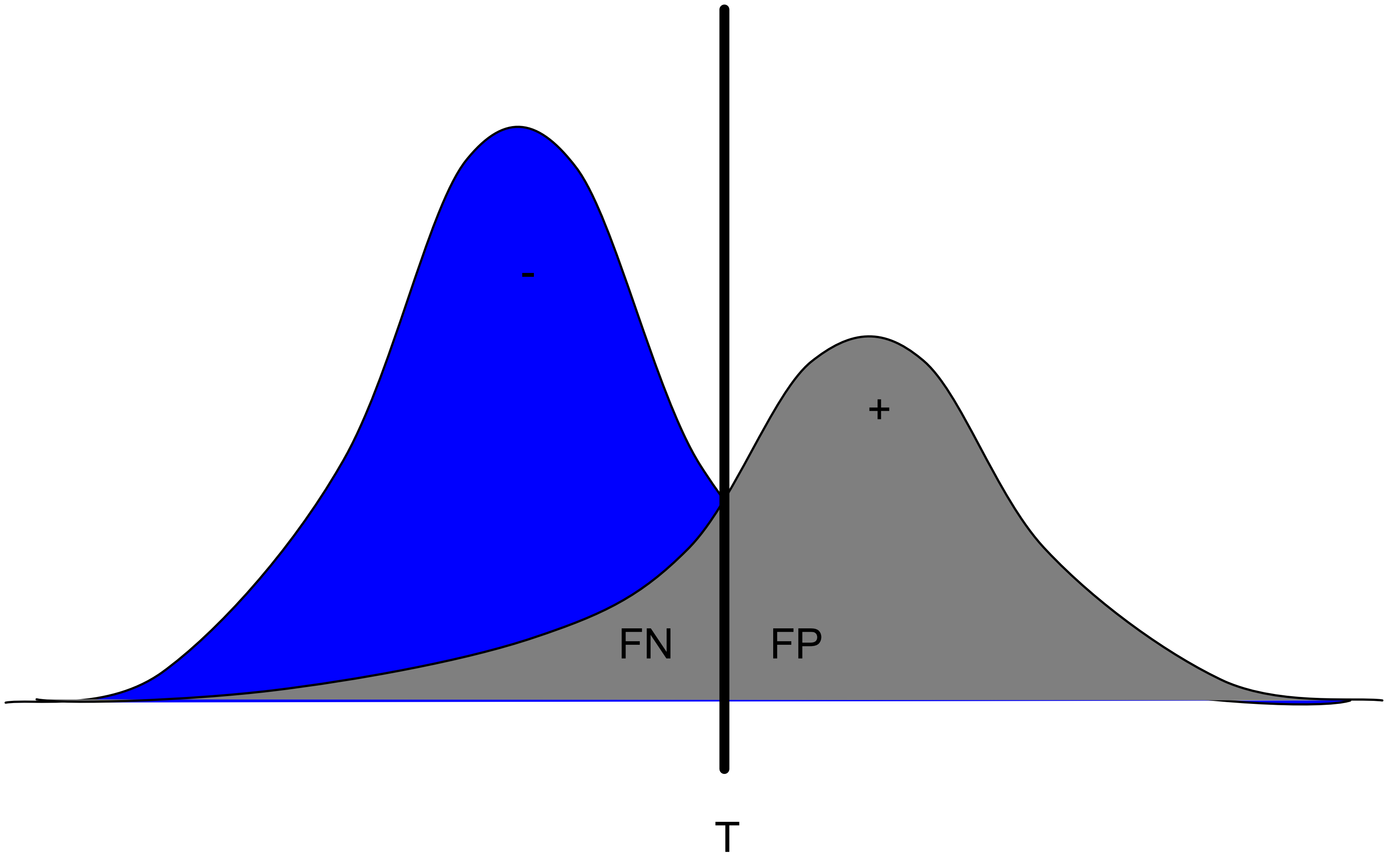
Specificity (true negative rate): TN/(FP+TN)

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

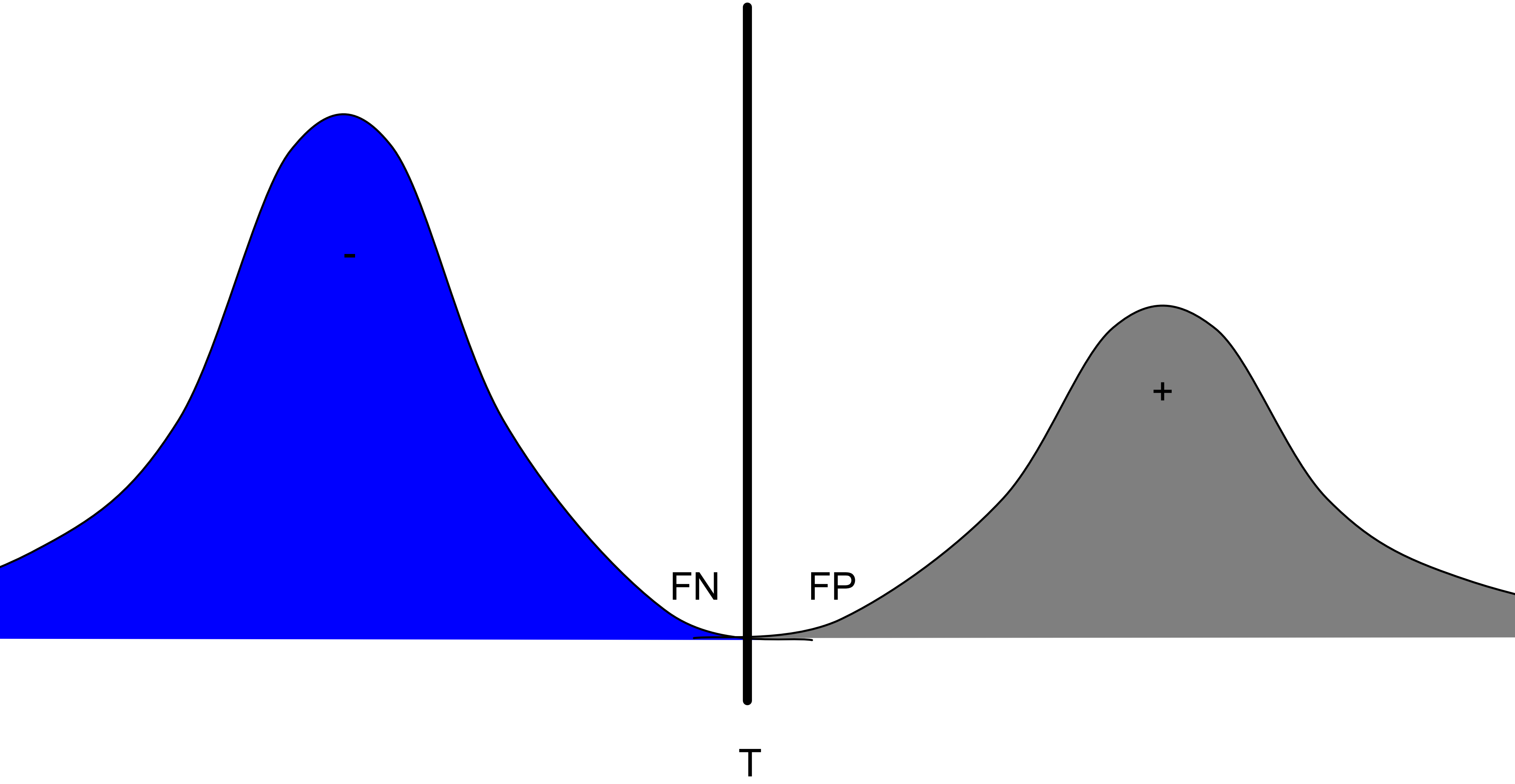
Positive Predictive Value: TP/(TP+FP)

Negative Predictive Value: TN/(FN+TN)

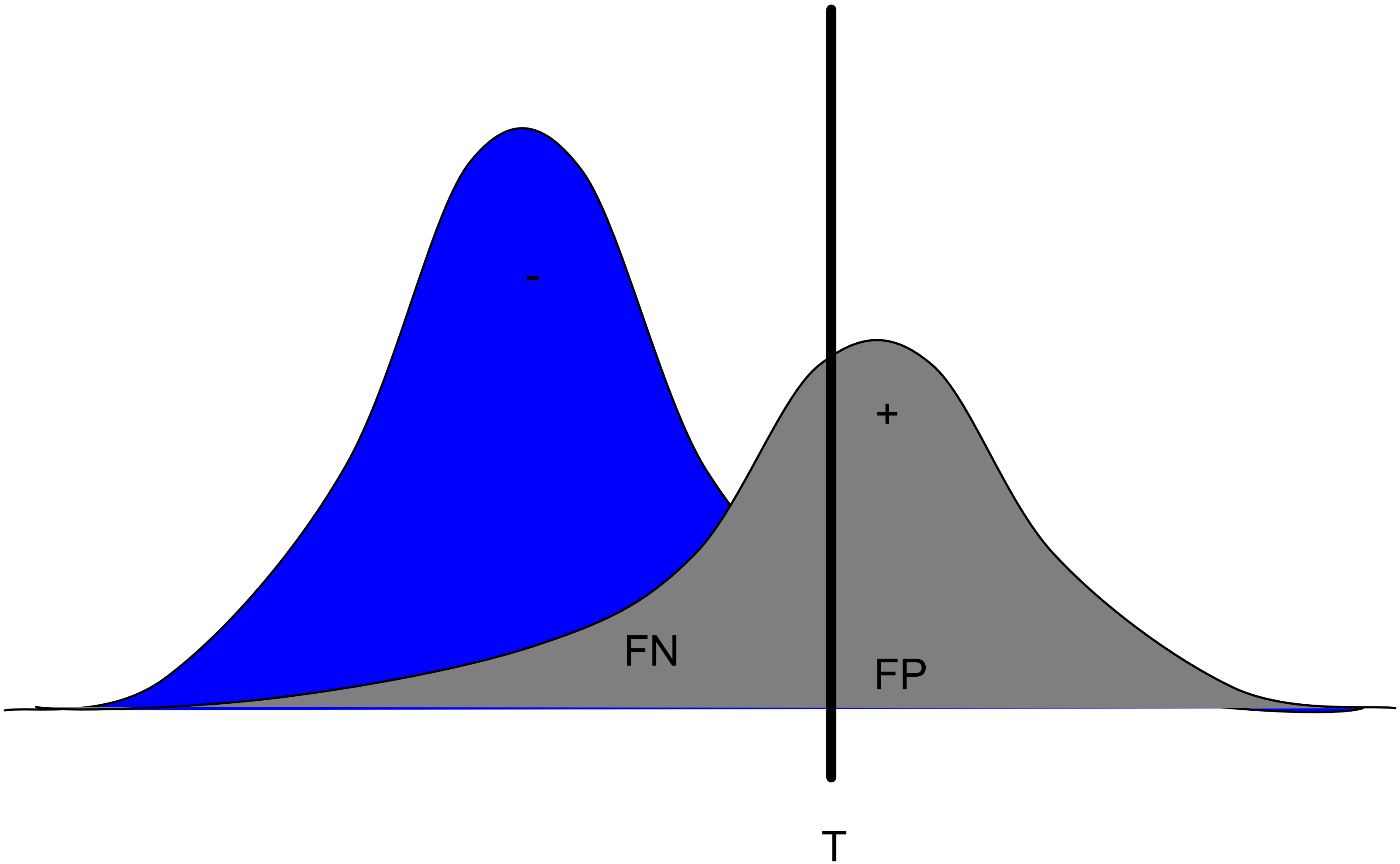
Test Thresholds



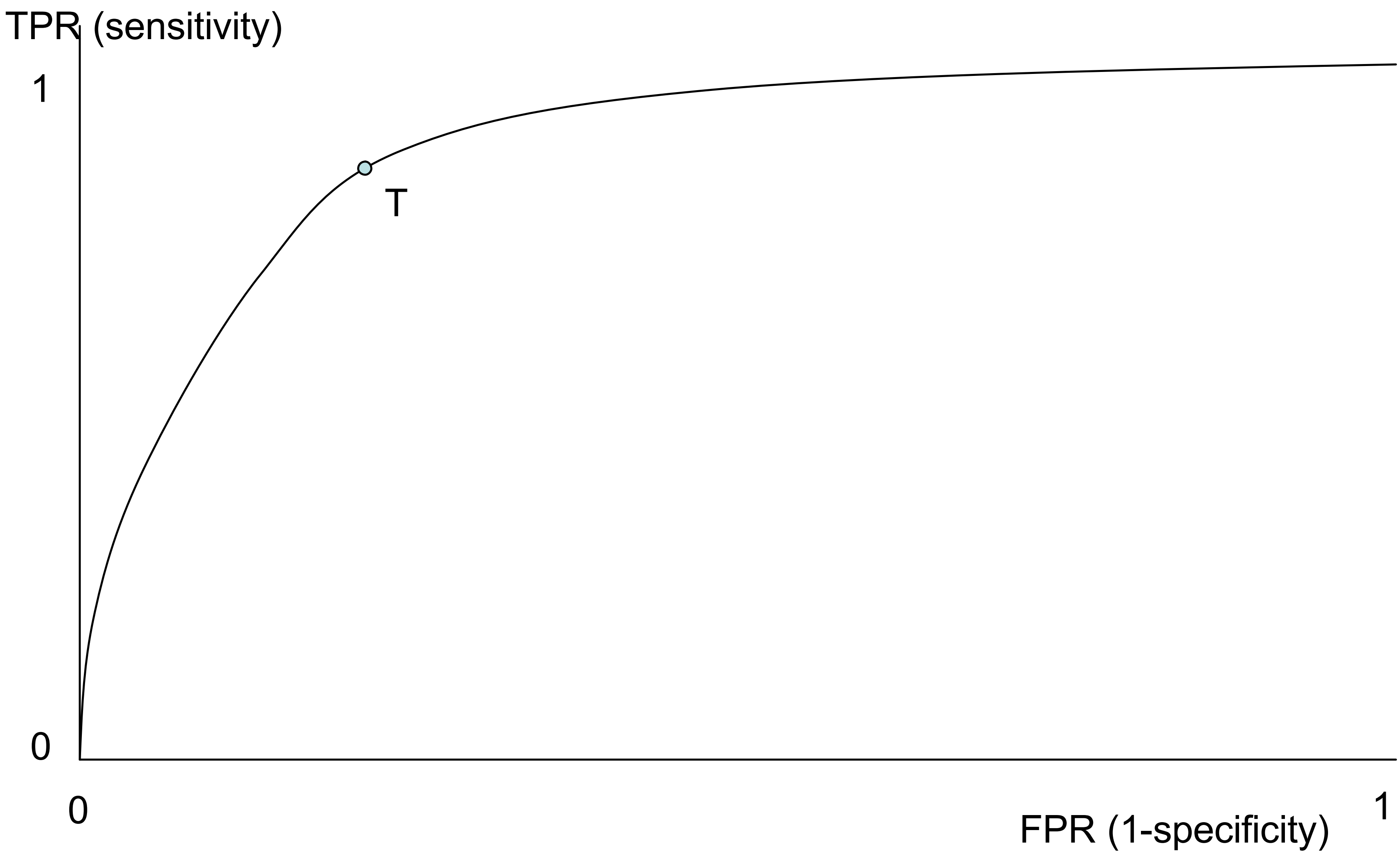
Wonderful Test



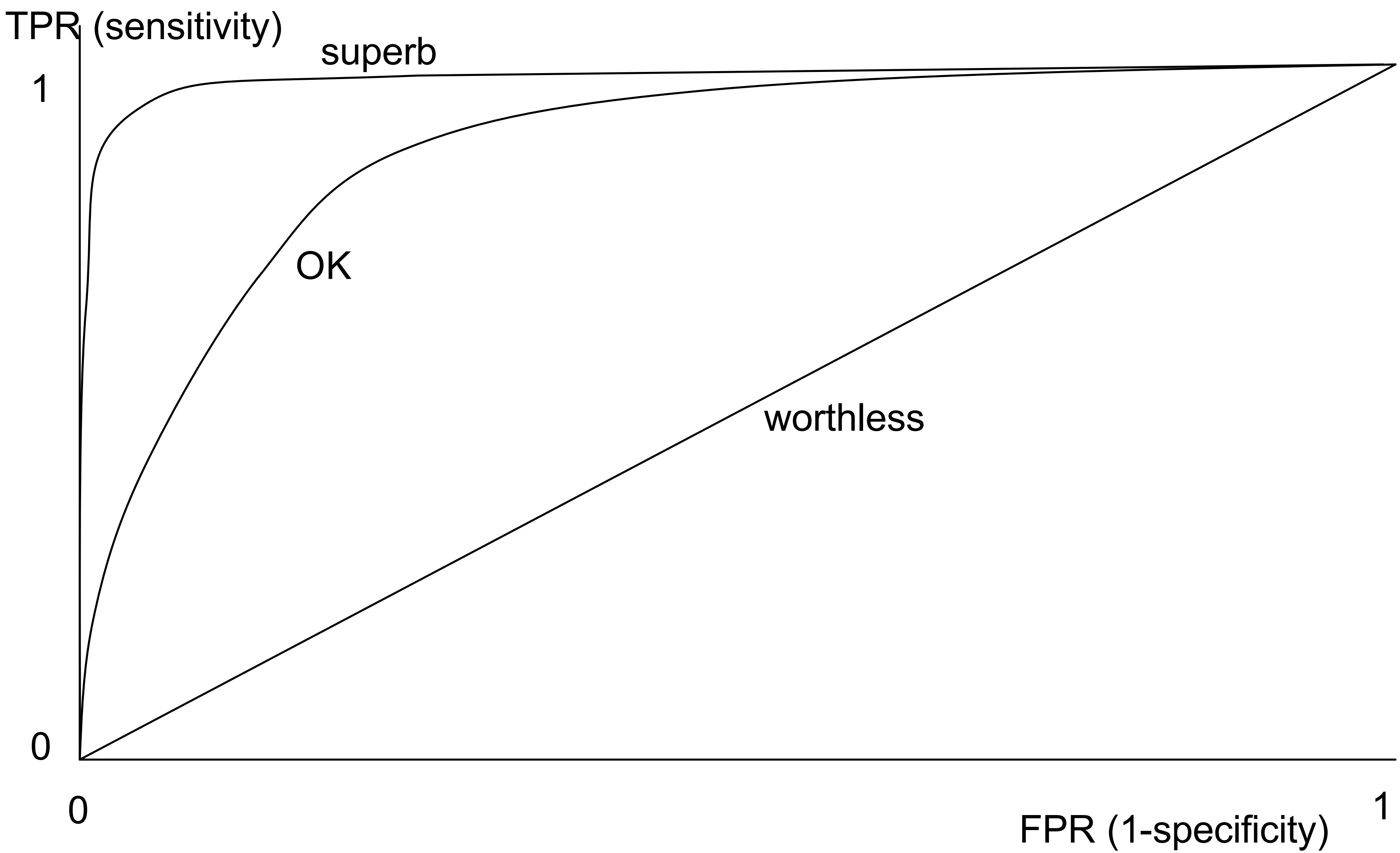
Test Thresholds Change Trade-off between Sensitivity and Specificity



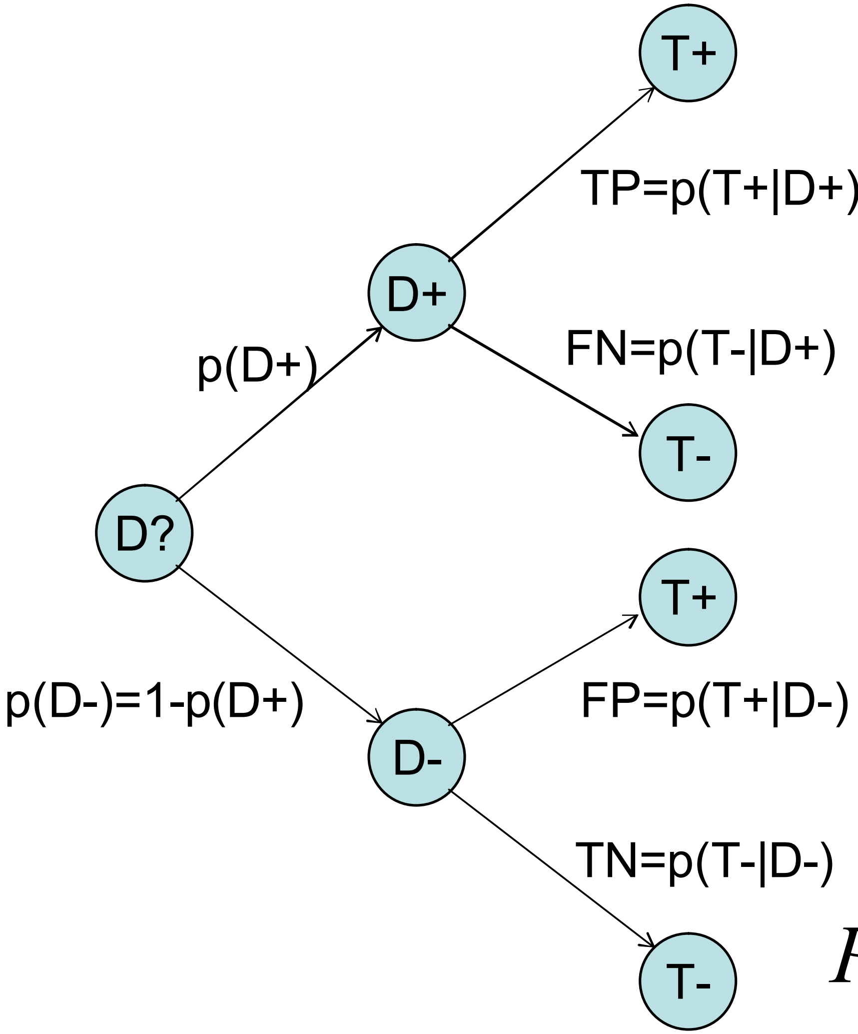
Receiver Operator Characteristic (ROC) Curve



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

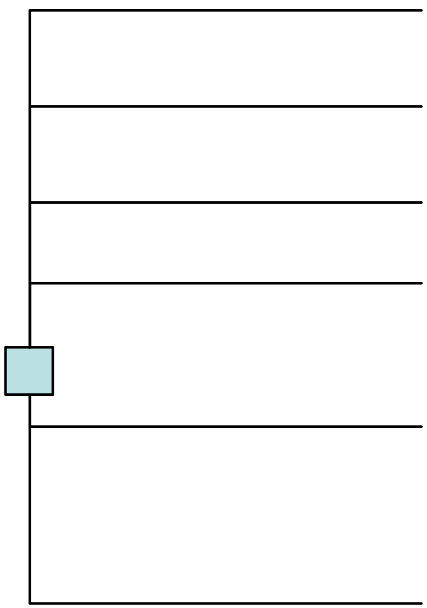
Example: Acute Renal Failure

- 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

Decision Tree for ARF

- Choose:
  - Surgery for obstruction
  - Treat with antibiotics
  - Perform pyelogram
  - Perform arteriography
  - Measure patient's temperature
  - Determine if there is proteinuria
  - …

Decision Tree for ARF



Surgery for obstruction

Value = ???

Treat with antibiotics

Perform pyelogram

Perform arteriography

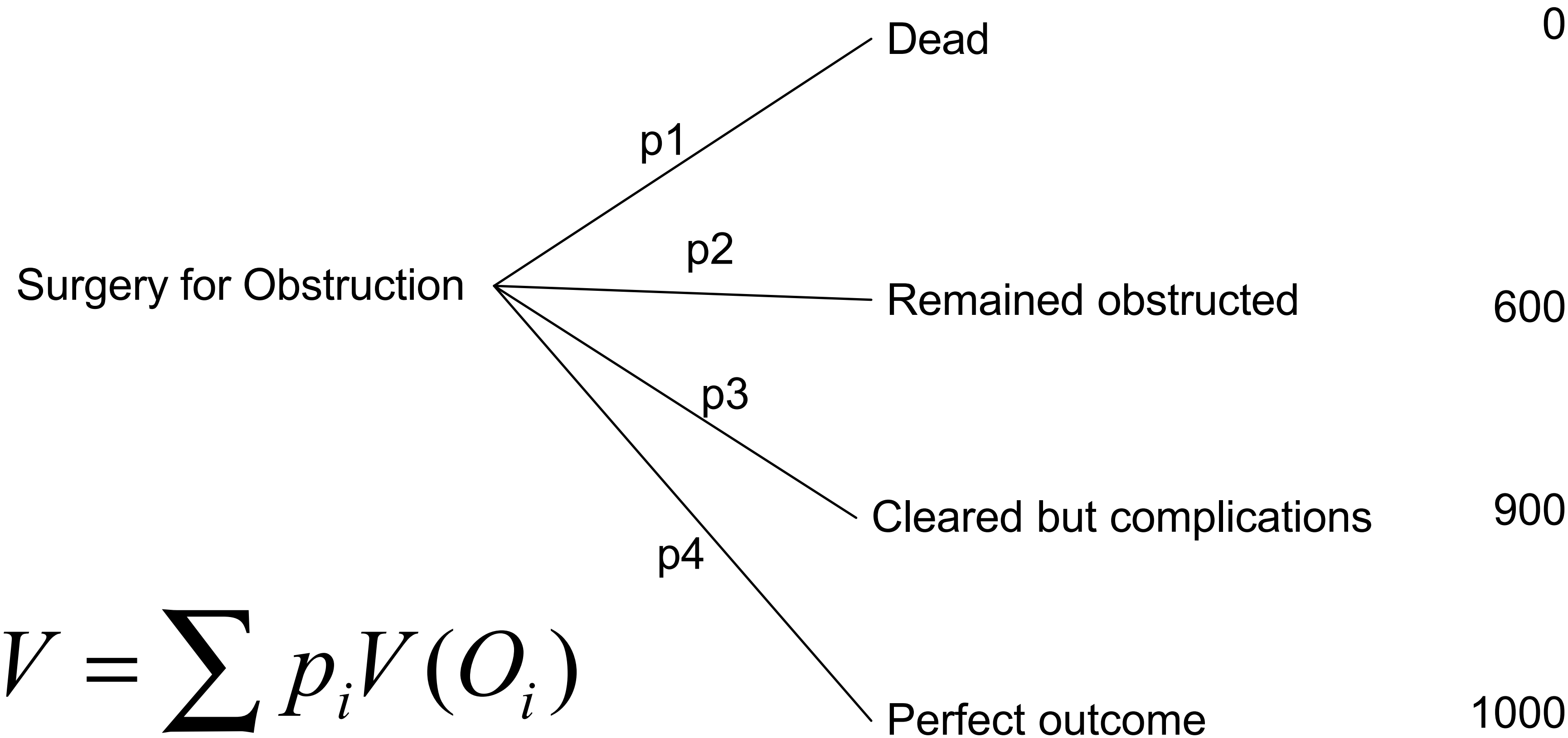
Measure patient’s temperature

Determine if there is proteinuria

What happens when we act?

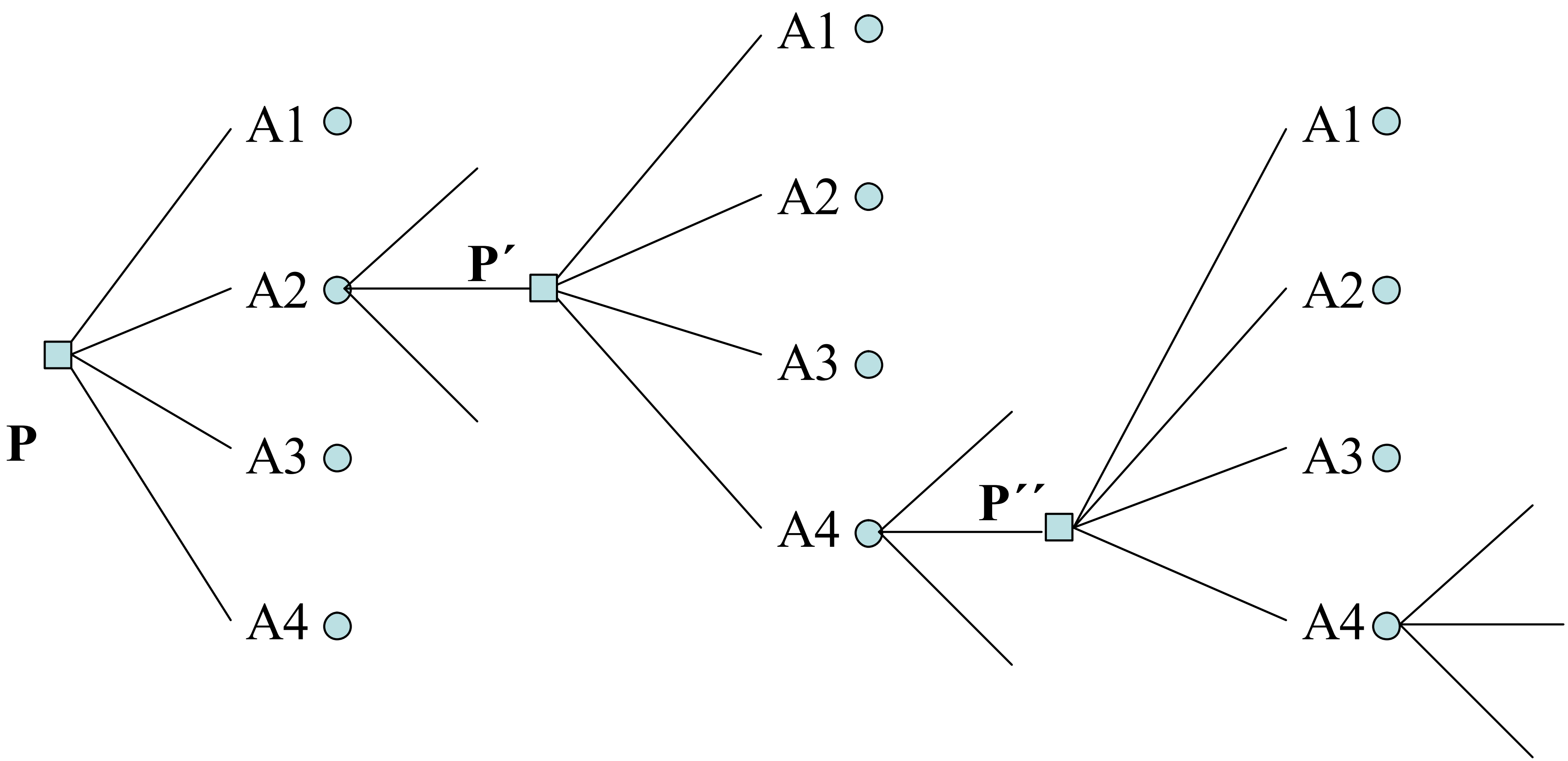
- Treatment: leads to few possible outcomes
  - different outcomes have different probabilities
    - probabilities depend on distribution of disease probabilities
  - value of outcome can be directly determined
    - value may depend on how we got there (see below)
    - therefore, value of a treatment can be determined by expectation
- Test: lead to few results, revise probability distribution of diseases, and impose disutility
- Questions: lead to few results, revise probability distribution

Treatment Outcome
(not as in ARF)



$$V = \sum_i p_i V(O_i)$$

Full decision tree



Initial probability distribution

ATN	**Acute tubular necrosis**	**0.250**
FARF	**Functional acute renal failure**	**0.400**
OBSTR	**Urinary tract obstruction**	**0.100**
AGN	**Acute glomerulonephritis**	**0.100**
CN	**Renal cortical necrosis**	**0.020**
HS	**Hepatorenal syndrome**	**0.005**
PYE	**Pyelonephritis**	**0.010**
AE	**Atheromatous Emboli**	**0.003**
RI	**Renal infarction (bilateral)**	**0.002**
RVT	**Renal vein thrombosis**	**0.002**
VASC	**Renal vasculitis**	**0.050**
SCL	**Scleroderma**	**0.002**
CGAE	**Chronic glomerulonephritis, acute exacerbation**	**0.030**
MH	**Malignant hypertension & nephrosclerosis**	**0.030**

ARF's Database: P(obs|D)

Conditional probabilities for <u>Proteinuria</u> **Diseases**	**Probabilities** **0**	**Trace to 2+**	**3+ to 4+**
ATN	0.1	0.8	0.1
FARF	0.8	0.2	0.001
OBSTR	0.7	0.3	0.001
AGN	0.01	0.2	0.8
CN	0.01	0.8	0.2
HS	0.8	0.2	0.001
PYE	0.4	0.6	0.001
AE	0.1	0.8	0.1
RI	0.1	0.7	0.2
RVT	0.001	0.1	0.9
VASC	0.01	0.2	0.8
SCL	0.1	0.4	0.5
CGAE	0.001	0.2	0.8
MH	0.001	0.4	0.6

Questions

- Blood pressure at onset
- proteinuria
- casts in urine sediment
- hematuria
- history of prolonged hypotension
- urine specific gravity
- large fluid loss preceding onset
- kidney size
- urine sodium
- strep infection within three weeks
- urine volume
- recent surgery or trauma
- age
- papilledema
- flank pain
- skin, intestinal or lung lesions
- history of proteinuria
- symptoms of bladder obstruction
- exposure to nephrotoxic drugs
- disturbance in clotting mechanism
- pyuria
- bacteriuria
- sex
- transfusion within one day
- jaundice or ascites
- ischemia of extremities or aortic aneurism
- atrial fibrillation or recent MI

Invasive tests and treatments

- Tests
  - biopsy
  - retrograde pyelography
  - transfemoral arteriography
- Treatments
  - steroids
  - conservative therapy
  - iv-fluids
  - surgery for urinary tract obstruction
  - antibiotics
  - surgery for clot in renal vessels
  - antihypertensive drugs
  - heparin

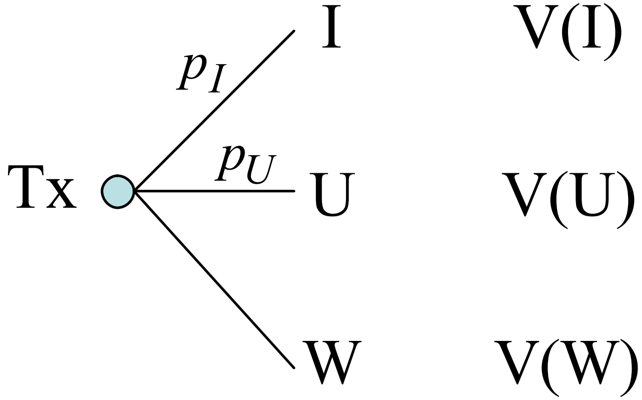
Updating probability distribution

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

Bayes' rule

Value of treatment

- Three results: improved, unchanged, worsened
  - each has an innate value, modified by “tolls” paid on the way
- Probabilities depend on underlying disease probability distribution



Modeling treatment

		Steroids	
	improved	*unchanged*	*worse*
atn	0.60	0.20	0.20
farf	0.05	0.35	0.60
obstr	0.05	0.60	0.35
agn	0.40	0.40	0.20
cn	0.05	0.75	0.20
hs	0.05	0.05	0.90
pye	0.05	0.05	0.90
ae	0.05	0.70	0.25
ri	0.01	0.14	0.85
rvt	0.10	0.30	0.60
vasc	0.15	0.25	0.60
scl	0.05	0.05	0.90
cgae	0.40	0.35	0.25
mh	0.05	0.05	0.90

Utilities:

improved: 5000

unchanged: -2500

worse: -5000

Modeling test: transfemoral arteriography

	p(clot)	*cost*
atn	0.01	500
farf	0.01	800
obstr	0.01	500
agn	0.01	500
cn	0.01	500
hs	0.01	800
pye	0.01	500
ae	0.03	800
ri	0.85	500
rvt	0.50	500
vasc	0.01	500
scl	0.01	500
cgae	0.01	500
mh	0.01	500

How large is the tree?

- Infinite, or at least (27+3+8)^(27+3+8), ~10^60
- What can we do?
  - Assume any action is done only once
  - Order:
    - questions
    - tests
    - treatments
- 27! x 4 x 3 x 2 x 8, ~10^30
- Search, with a *myopic evaluation function*
  - like game-tree search; what's the static evaluator?
  - Measure of certainty in the probability distribution

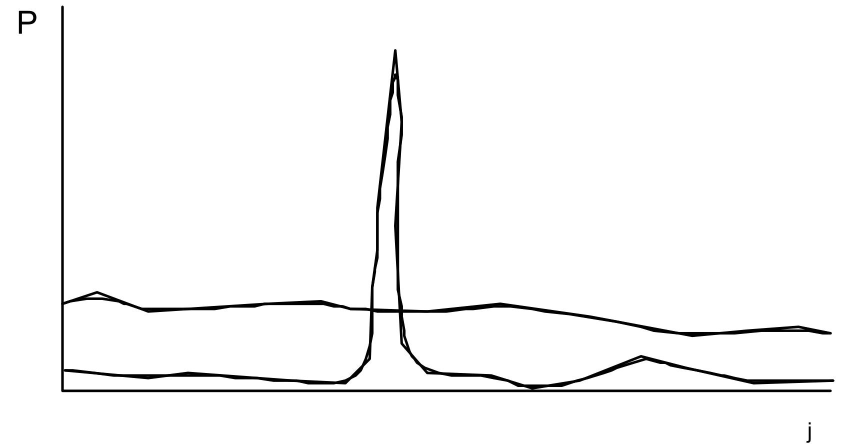
How many questions needed?

- How many items can you distinguish by asking 20 (binary) questions? 2^20
- How many questions do you need to ask to distinguish among *n* items? $\log_2(n)$
- *Entropy* of a probability distribution is a measure of how certainly the distribution identifies a single answer; or how many more questions are needed to identify it

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
. . .

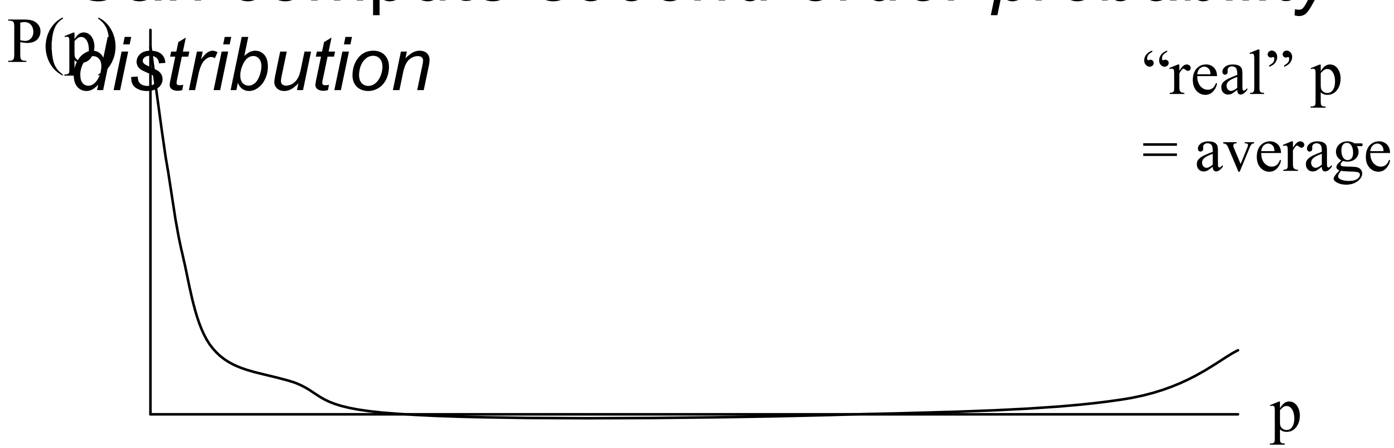
Local Sensitivity Analysis

Case-specific Likelihood Ratios

Therapy Planning Based on Utilities

Global Sensitivity Analysis

- When asking questions, “how bad could it get for the leading hypothesis?”
  - Assume all future answers are worst possible in terms of likelihood ratio P(obs|D)/P(obs|~D)
  - Usually, (0,1)
  - Can compute *second order probability distribution*



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*

Demo of ARF & Similar Programs