A.I. in health informatics **lecture 2** clinical reasoning & probabilistic inference, I

kevin small &
byron wallace

today

- a review of probability
  - random variables, maximum likelihood, etc.
  - crucial for clinical reasoning ***and*** AI
  - … and also AI-clinical reasoning systems

- intro to (review of?) 2x2 tables, basic probability, counting, etc.

probability (& statistics)

- quantify uncertainty
  - both from measurement and sampling errors
- *how likely is it that an event e will occur*?
  - at the end of the day: counting
- let's consider an example…

what are the chances?

- $p(X)$ denotes the probability of event X occurring
- usually interpreted as: the fraction of outcomes (out of all possible outcomes) in which X occurs
  - philosophically loaded, but that's not what this class is about …

probability: some basics

- sum rule

$$p(X) = \sum_Y p(X,Y)$$

“Probability of Y given X”

- product rule

$$p(X,Y) = p(Y \mid X)\, p(X)$$

- Bayes’ theorem

$$p(Y \mid X) = \frac{p(X \mid Y)\, p(Y)}{p(X)}$$

more on this later!

random variables

- *random variables* follow some latent distribution
  - we usually assume we know its form, but not its parameters
  - this is the *generative story*

expectations

- E[X] denotes the *expectation* of a random variable X

- it is the average of the values it can take, weighted by their respective probabilities of occurring

$$\mathrm{E}[X] = \sum p(x) \cdot x$$

discrete, finite case

$$\mathrm{E}[X] = \int p(x) \cdot x$$

continuous case

expectations

- expectations are vital for formal reasoning

- should I play the lottery?
  - perhaps if the expected value of lottery ticket is >0
  - $E[X] = (-\$1)*(.9999)... + (+\$10000)*(.0…01)$
  - probably not

probability: balls & bins

- suppose we are (inexplicably!) given a jar filled with red and blue balls
- we draw n balls* – m are blue
- what proportion of the balls in the jar are blue?

* *with replacement*

probability: balls & bins

- *m/n* (observed proportion) is the ‘best’ (*maximum likelihood)* estimate of *p*
  - ML estimates usually wear hats $\hat{p}$

- … wait, wh e best?

(it’s not always.)

maximum likelihood

the maximum likelihood approach: given observed data D, pick the parameters **w** that maximize the probability of D, p(D|**w**)

balls & bins, ml style

- let p denote the probability of drawing a blue ball (equiv., proportion of blue balls)

- probability of a draw being: blue = p; red =1-p

- probability of an observed m blue balls in a sequence of N draws

$$p^m (1-p)^{N-m}$$

balls & bins, ml style

- thus the likelihood function *L* is $p^m(1-p)^{N-m}$
- if you take the derivative w.r.t. *p* and set it to zero (maximizing the likelihood function), you get: $p=m/N$
- note that this ignores prior/external information!

a word on Bayes

- suppose we flip a coin that we believe to be fair three times, and it comes up heads each time
  - the ML estimate will imply that the coin will *always* land on heads
  - but incorporating prior beliefs can avoid this

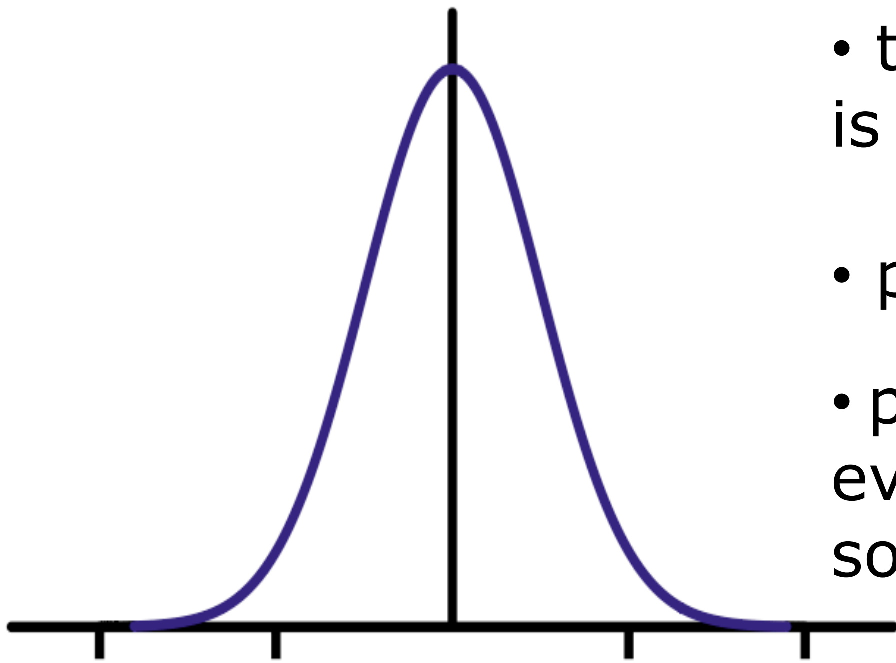
- the Bayesian way:

posterior α likelihood x prior

this example in bishop, 2006

the Gaussian

- we used the Bernoulli (for binary events) is what we used in the balls & jar example



- the Gaussian (a/k/a normal) is ***the*** distribution
- parameterized by $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$
- people typically just assume everything is normal. sometimes wrongly…

many other distributions!

- Poisson – p(# of events occurring in an interval)
- Beta – often used as a prior in Bayesian analysis
- Multinomial – generalization of binomial (which we saw before)

the 2x2 table

- confusingly called the ‘confusion matrix’

reference standard (‘truth’)

predictor (test, classifier…)

	positive	negative
positive	true positive	false positive
negative	false negative	true negative

common metrics

reference standard ('truth')

predictor (test, classifier...)

	positive	negative
positive	true positive	...lse positive
negative	false ne...	...negative

people misinterpret these all the time! see homework...

sensitivity a/k/a recall $= \dfrac{tp}{tp + fn}$

tn

fp

precision $= \dfrac{tp}{tp + fp}$

accuracy $= \dfrac{tp + tn}{tp + fp + tn + fn}$

choosing between tests

- recall:

sensitivity $\dfrac{tp}{tp + fn}$ specificity $\dfrac{tn}{tn + fp}$

- but these depend on a cut-off!

- how to assess test performance independent of this threshold?

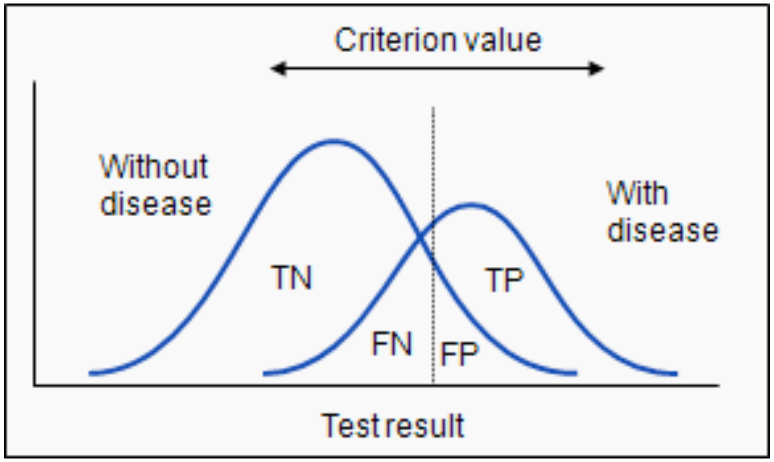
ROC-space

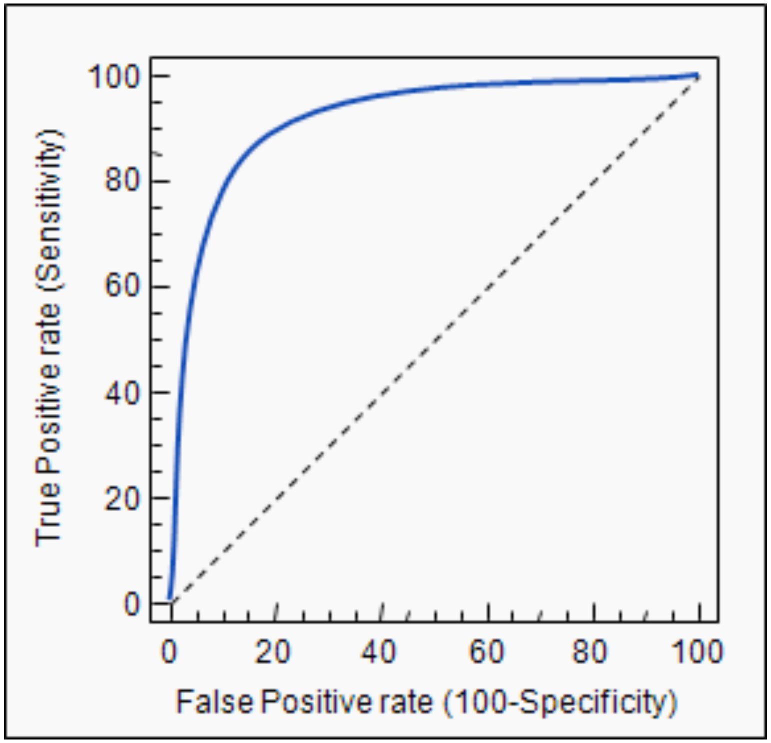
- ***ROC Receiver Operating Characteristic*** (ROC) curve

- ROC space is 2-d: *sensitivity* and *1-specificity*

- ideal tests occupy upper right region

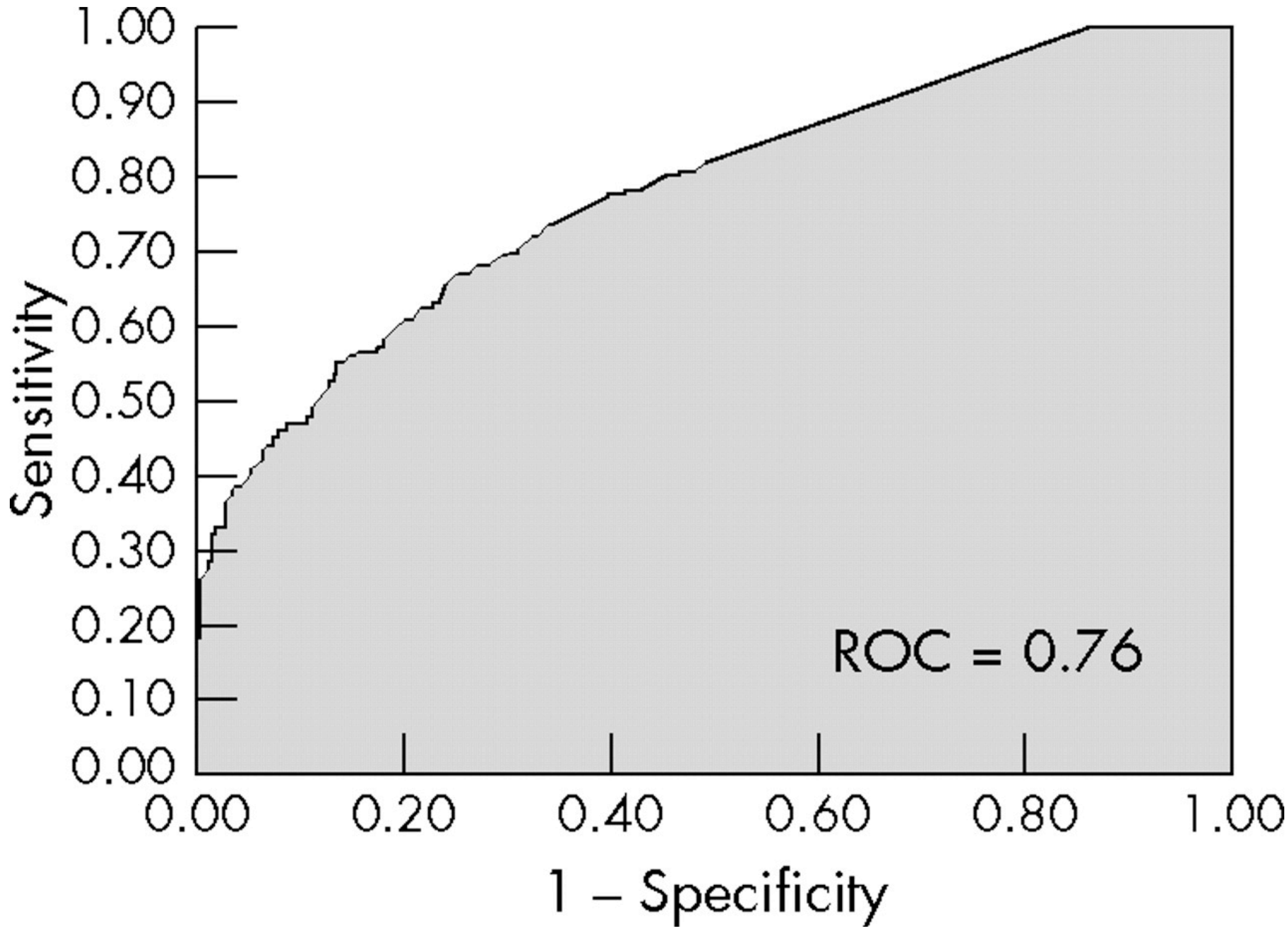
ROC-space





images from http://www.medcalc.org/manual/roc-curves.php

AUC



1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Sensitivity
ROC = 0.76
0.00
0.20
0.40
0.60
0.80
1.00
1 – Specificity

uncommon metrics

reference standard ('truth')

predictor (test, classifier...)

	positive	negative
positive	true positive	false positive
negative	false negative	true negative

negative predictive value (npv) = $$\frac{tn}{tn + fn}$$

likelihood ratio (+) = $$\frac{sensitivity}{1 - specificity}$$

likelihood ratio (-) = $$\frac{specificity}{1 - sensitivity}$$

humans are not good at probability

People are [Tversky & Hahneman]:

Whoops, forgot about this guy

- *insensitive to priors* – p(y)p(y|x)

- insensitive to *sample size* – the probability that the average in a sample is more extreme than the population average increases with small samples

- *Gambler's fallacy*

misleading intuitions

true hiv status

hiv test

	positive	negative
positive	true positive	false positive
negative	false negative	true negative

- assume test has 98% sensitivity and 99% specificity
- … great test, right?

misleading intuitions

- HIV prevalence in USA = ~.3%
- suppose someone tests positive; what's the probability that they have HIV?
- X =^{def} event of a positive test result; Y =^{def} event that the patient actually has HIV; ~Y =^{def} does not have HIV

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X \mid Y)P(Y) + P(X \mid \sim Y)P(\sim Y)}$$

$$= \frac{.99 \cdot .003}{.99 \cdot .003 + .01 \cdot .997} \approx .23$$

misleading intuitions

- so: what is the likelihood that the a person with positive test results has HIV?
  - 23% -- **it's 3x more likely that they *don't* have HIV, despite a positive test result!**

$$P(\sim Y \mid X) = \frac{P \ldots}{P(X \mid \sim \ldots}$$

$$= \frac{.01 \cdot .99\ldots}{.01 \cdot .997 + .9\ldots .003}$$

classic example of insensitivity to priors!

automatic diagnosis

- assume we have a database of diseases, their prevalences and test characteristics.

- really naïve automated dia m:

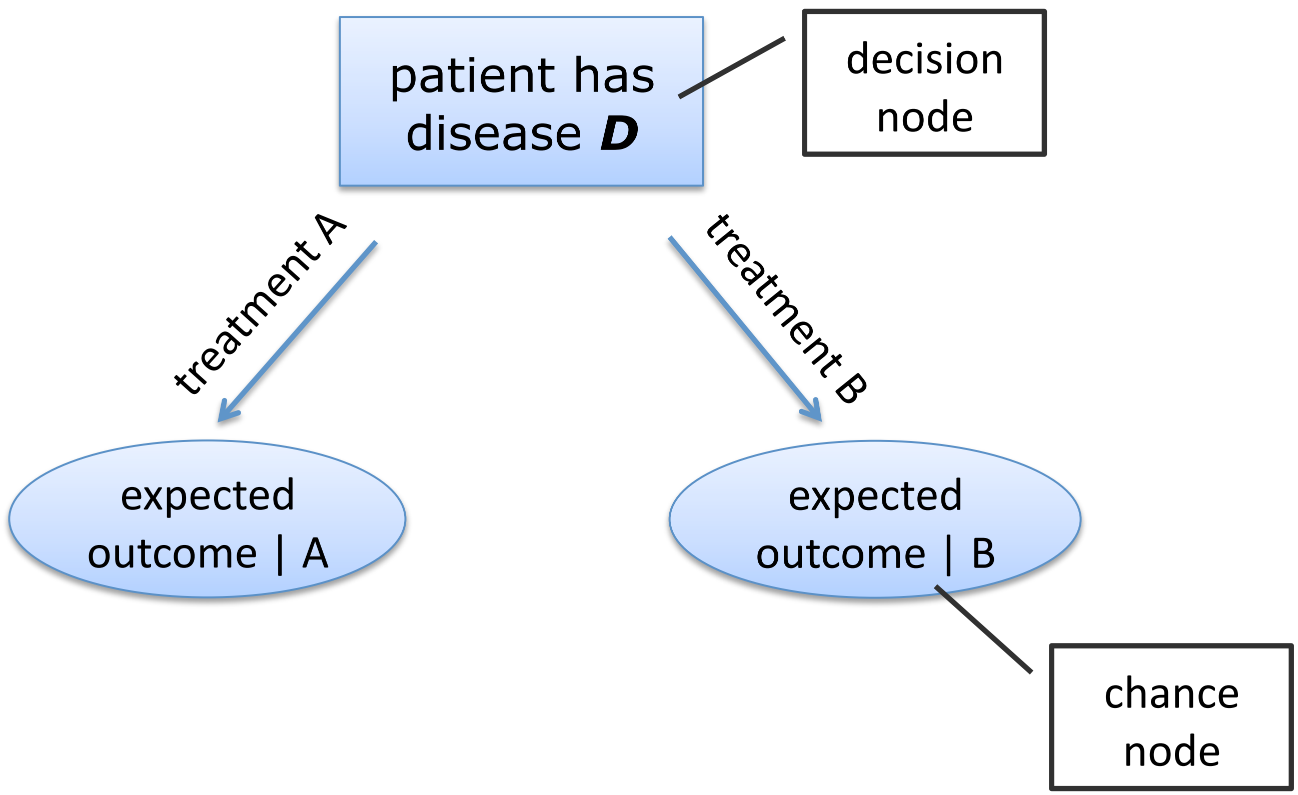
```
diagnose (test result X)
  return arg max_Y P(X|Y)P(Y) ses Y
```

this is how the Leeds abdominal pain system worked!

a calculus for decisions

- need to integrate *utilities* and *costs* into a probabilistic framework
- next up: ***decision theory***

expected-value decision making



the outcomes are stochastic, so we depend on the *expected outcome*

expectations (review)

- E[X] denotes the *expectation* of a random variable X

- it is the average of the values it can take, weighted by their respective probabilities of occurring

$$\mathrm{E}[X] = \sum p(x) \cdot x$$

discrete, finite case

$$\mathrm{E}[X] = \int p(x) \cdot x$$

continuous case

decision making made easy

1. calculate E[x] over all decision alternatives x
2. select the decision x* that maximizes E[x]

but …

- creating the decision tree is non-trivial
  - we’ll come back to this

- we’ve not taken into account variance of the expectation, nor values of outcomes, nor costs

utilities

- encode arbitrary (potentially subjective) ‘pay-offs’
  - by convention, between 0 and 1
  - perfect health = 1; death = 0

- where do utilities come from?

the standard gamble

[von neumann & morgenstern]

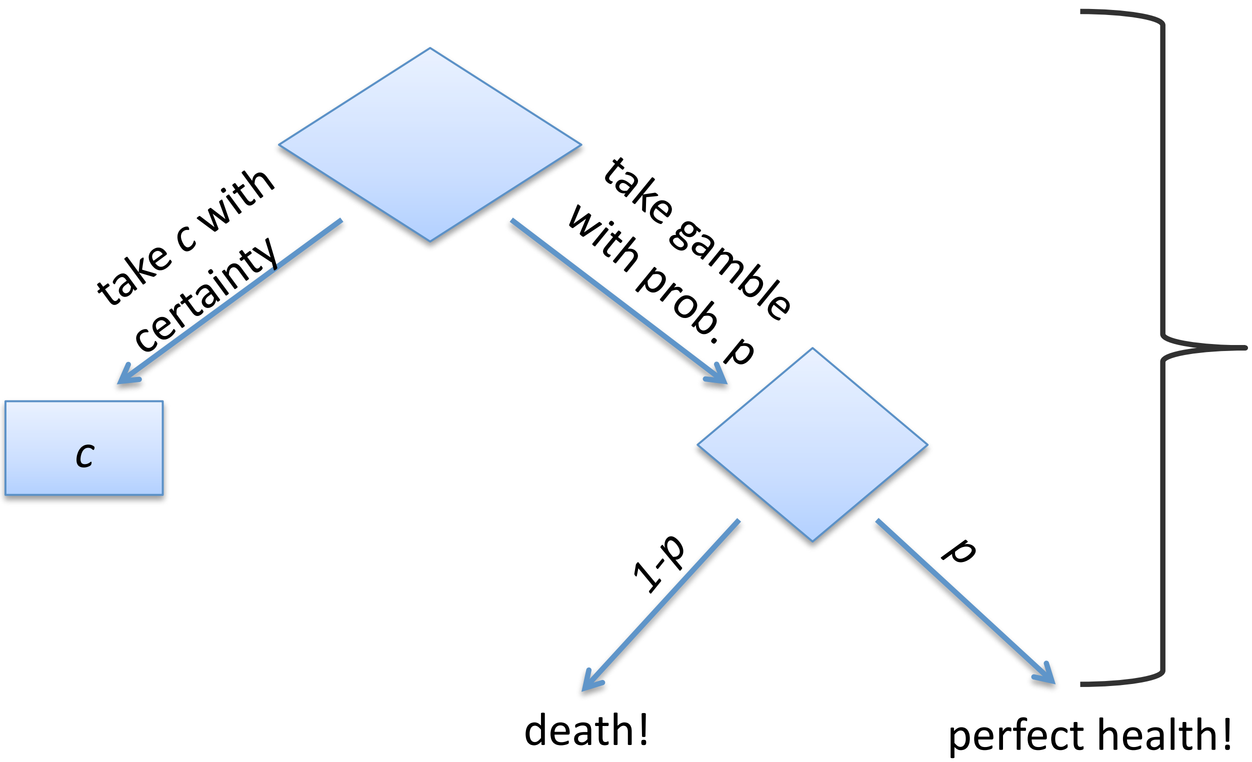
- the (morbid) gamble: we will flip a coin with bias *p* – tails implies death (0); heads perfect health (1)

- let *c* $=^{\text{def}}$ some relevant clinical state (e.g. paralysis)

- the *utility* of *c* is the value *p** at which the patient has no preference between:
    1) *c* (being paralyzed)
    2) taking the gamble

the standard gamble

[von neumann & morgenstern]



at what value of *p* is the patient indifferent to which path is taken?

time trade-off technique

[sox et a.]

- ask the patient to decide the length of life t in perfect health equivalent to a given length of life with c, t^c
  - e.g., “I find 6 months of life paralyzed to be equivalent to 1 year of life with full mobility”

- utility = t^c/t
  - above: 6/12 = .5

utility-maximization

- associate chance nodes with utilities

```
select-treatment ()
  return arg max_T Expected Utility[T]
```

- expected utility is defined recursively down branches (until terminal nodes are reached)