Mathematics for Computer Science
MIT 6.042J/18.062J

Testing with Confidence

98% accurate TB testing

A great-sounding diagnostic test for TB:

98% accurate TB testing

A great-sounding diagnostic test for TB: if someone has TB the test is guaranteed to detect it.

98% accurate TB testing

A great-sounding diagnostic test for TB: if someone has TB the test is guaranteed to detect it. If they don't have TB, the test says so 98% of the time.

98% accurate TB testing

$$\Pr[\;+\;\mid TB] = 1$$

"+" for [test positive]

98% accurate TB testing

$$\Pr[\;+\;\mid TB] = 1$$

$$\Pr[\,+\mid not(TB)] = \frac{2}{100}$$

(false positive rate only 2%)

98% accurate TB testing

Overall

$$\Pr[\text{ mistake}] < \frac{1}{50}$$

98% accurate TB testing

Your doctor tests you, and it says TB!

98% accurate TB testing

Your doctor tests you, and it says TB! He says "The hypothesis that you have TB holds at the 98% confidence level."

98% accurate TB testing

TB is a serious disease, and your Doc is "98% confident" you have it.

98% accurate TB testing

TB is a serious disease, and your Doc is "98% confident" you have it. Should you get treatment?

Do you have TB?

TB is a serious disease, and your Doc is "98% confident" you have it. Should you get treatment?

Do you have TB?

TB is a serious disease, and your Doc is "98% confident" you have it. Should you get treatment? …depends on probability you have TB.

A Random Person

We don't really mean the "probability" that you personally have TB. We're just thinking of you as a random person.

Do you have TB?

You know 2%

Pr[+ | no TB].

Do you have TB?

Confidence indicates

Pr[+ | no TB].

Confidence vs Prediction

Confidence indicates

$\Pr[+ \mid \text{no TB}]$.

You want prediction:

Confidence vs Prediction

Confidence indicates

$\Pr[+ \mid \text{no TB}]$.

You want prediction:

$\Pr[\text{no TB} \mid +]$.

Do not confuse these!

Confidence vs Prediction

Bayes' Theorem lets us

find $\Pr[\text{noTB} \mid +]$

given $\Pr[+ \mid \text{noTB}]$

Odds of an Event

Event E

$$\text{Odds}[E] ::= \frac{\Pr[E]}{\Pr[\overline{E}]}$$

Odds of an Event

example: 6-sided die

$$\Pr[\text{roll } 3] = \frac{1}{6}$$

$$\text{Odds}[\text{roll } 3] = \frac{1/6}{5/6} = \frac{1}{5}$$

"1 to 5"

Odds of TB

$$\text{Odds}[\text{no TB}] = \frac{\Pr[\text{no TB}]}{\Pr[\text{TB}]}$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +]$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +] = \frac{\Pr[\text{no TB} \mid +]}{\Pr[\text{TB} \mid +]}$$

$$= \frac{\Pr[+ \mid \text{no TB}]\Pr[\text{no TB}] / \Pr[+]}{\Pr[+ \mid \text{TB}]\Pr[\text{TB}] / \Pr[+]}$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +] = \frac{\Pr[\text{no TB} \mid +]}{\Pr[\text{TB} \mid +]}$$

$$= \left(\frac{\Pr[+ \mid \text{no TB}]}{\Pr[+ \mid \text{TB}]}\right)\left(\frac{\Pr[\text{no TB}] \,/\, \cancel{\Pr[+]}}{\Pr[\text{TB}] \,/\, \cancel{\Pr[+]}}\right)$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +] =$$

$$\frac{\Pr[+ \mid \text{no TB}]}{\Pr[+ \mid \text{TB}]} \cdot \frac{\Pr[\text{no TB}]}{\Pr[\text{TB}]}$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +] =$$

$$\underbrace{\frac{\Pr[+ \mid \text{no TB}]}{\Pr[+ \mid \text{TB}]}}_{\text{Bayes' factor}} \cdot \text{Odds}[\text{no TB}]$$

Do you have TB?

$$\text{Odds}[\text{no TB} \mid +] =$$

$$\underbrace{\frac{1/50}{1}}_{\text{Bayes' factor}} \cdot \text{Odds}[\text{no TB}]$$

Do you have TB?

$$\text{Odds[no TB} \mid +] = \underbrace{\frac{1}{50}}_{\text{Bayes' factor}} \cdot \text{Odds[no TB]}$$

11,000 TB cases reported

CDC got reports of 11,000 cases of TB in US in 2011.

Will be lots of unreported.

So estimate:

$$\Pr[\text{TB}] = \frac{1}{10,000}$$

11,000 TB cases reported

CDC got reports of 11,000 cases of TB in US in 2011.

Will be lots of unreported.

So estimate:

$$\text{Odds[no TB]} = 9,999$$

Do you have TB?

$$\text{Odds[no TB} \mid +] = \frac{1}{50} \cdot 9,999$$

Do you have TB?

$$\text{Odds[no TB} \mid +] = 199.98$$

$$\Pr[\text{no TB} \mid +] = 0.9950\ldots$$

Unlikely you have TB

Predicted probability of TB

$$\Pr[\text{TB} \mid +] \approx \frac{1}{200}$$

Unlikely you have TB

Because of relatively high false positive rate (2%) compared to TB rate (0.01%), predicted probability of TB remains small (1/2 %)!

A "more accurate" test

98% accurate test is not so good here.

A "more accurate" test

98% accurate test is not so good here. In fact, there's a trivial test that is 99.99% accurate:

always say "No TB"

98% accuracy still useful

98% accurate test did increase your odds of TB 50 times.

98% accuracy still useful

98% accurate test did increase your odds of TB 50 times. If you only had 7M medicine doses for a population of 350M, whom should you medicate?

98% accuracy still useful

If you medicate at random you'll only medicate

$$\frac{7}{350} = 2\%$$

of sick people.

98% accuracy still useful

Instead, medicate the 7M who test positive.

98% accuracy still useful

Instead, medicate the 7M who test positive. All the sick people are sure to be among these.