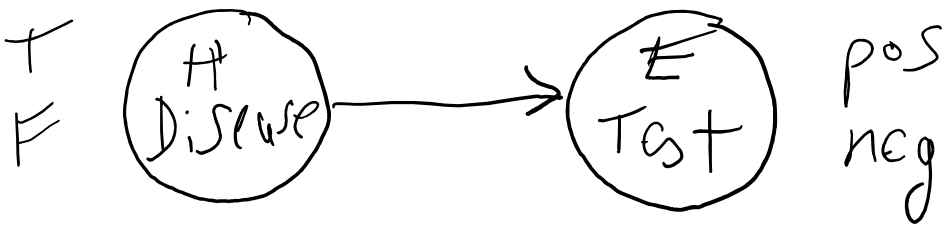
Week 2

Live Discussion Session

Starts at 2.05pm

Marginalisation



$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E)}$$

Shorthand

$$p(H=T \mid E=pos) = \frac{\overset{0.99}{p(E=pos \mid H=T)} \times \overset{0.05}{p(H=T)}}{p(E=pos)}$$

$p(H=T) = 0.001$

$p(E=pos \mid H=T) = 0.99$

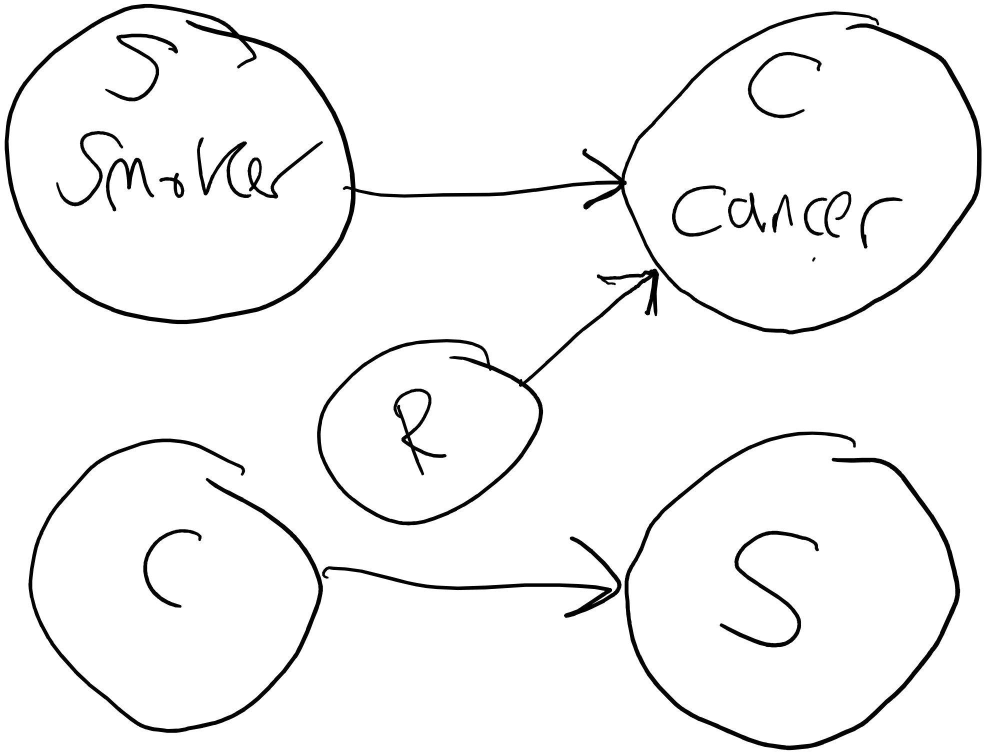
$p(E=pos \mid H=\cancel{T}F) = 0.\cancel{0}05$

$$p(E=pos) = p(E=pos \mid H=T) * p(H=T) + p(E=pos \mid H=F) * p(H=F)$$

$$= 0.99 \times 0.001 + 0.05 \times 0.999$$

$= 0.05094$

$= 5.09\%$



$p(C|S)$

$p(C|S, R)$

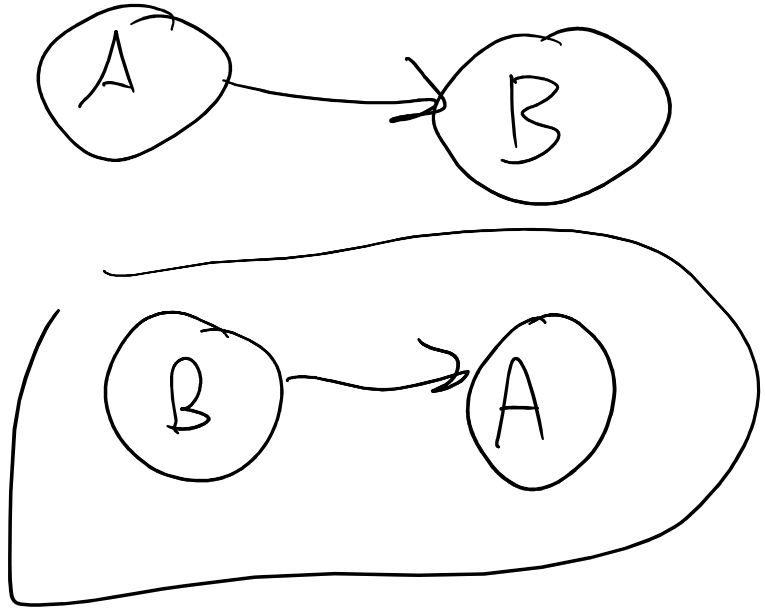
80% of those diagnosed with cancer are smokers

$p(S|C) = 0.8$

$p(C) = 0.$

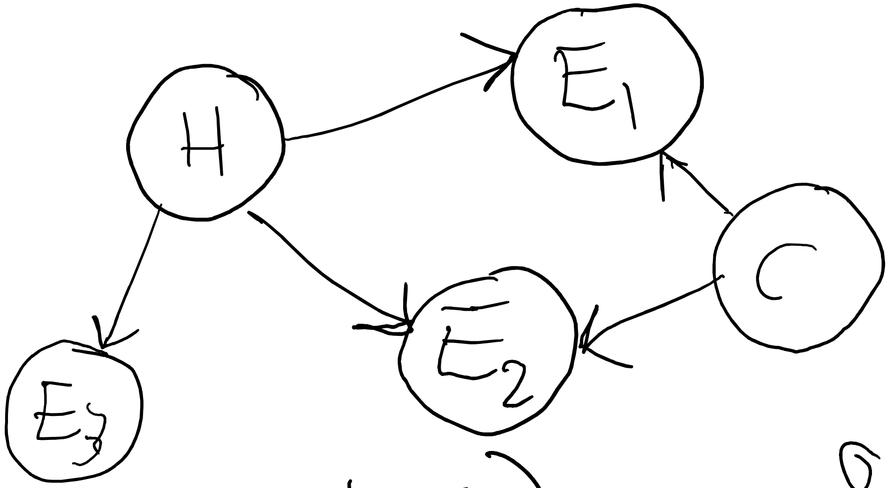
$p(S) = 0.5$

$$p(C|S) = \frac{p(S|C) \times p(C)}{p(S)} = 0.16$$



$p(B|A)$ ✓

$p(A|B)$



$$p(H \mid (E_1 \text{ and } \bar{E}_2)) = \frac{p((E_1 \text{ and } \bar{E}_2) \mid H) \times p(H)}{p(E_1 \text{ and } \bar{E}_2)}$$

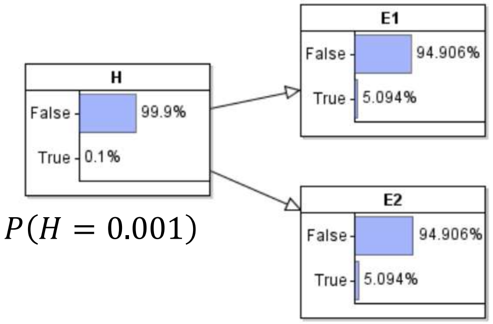
$$p(\bar{E}_2 \mid E_2) = \frac{\overset{0.99}{p(E_1 \mid H)} * \overset{0.99}{p(\bar{E}_2 \mid H)} \times \overset{0.001}{p(H)}}{\boxed{p(\bar{E}_1) * p(\bar{E}_2)}} = 0.282$$

0.05064^2

$= 28.2\%$

This is wrong. See next (corrected) slide

Here are our assumptions:



$P(H = 0.001)$

$P(E_1|H) = 0.99$ $\quad$ $P(E_1|H) = 0.05$

$P(E_2|H) = 0.99$ $\quad$ $P(E_2|H) = 0.05$

We want to calculate the probability of H if BOTH independent tests E1 and E2 are positive, i.e. calculate $P(H|(E1\ and\ E2))$

By Bayes theorem:

$$P(H|(E_1\ and\ E_2)) = \frac{P((E_1\ and\ E_2)|H) \times P(H)}{P(E_1\ and\ E_2)}$$

In the live lecture I said “Because E1 and E2 are independent we know that”:

$$P((E_1\ and\ E_2\ )|H) = \ P(E_1|H) \times P(E_2|H) \qquad (1)$$

$$P(E_1\ and\ E_2) = \ P(E_1) \times P(E_2) \qquad (2)$$

But E1 and E2 are ***only*** independent once we know whether H is true or false. So, while (1) is correct, (2) is NOT correct

But, by marginalisation, we known that: $P(E_1 \text{ and } E_2) \quad = P((E_1 \text{ and } E_2)|H) \times P(H) \quad + \quad P((E_1 \text{ and } E_2)|not\ H \ \times P(not\ H)$

$$= \ P(E_1|H) \times P(E_2|H) \times P(H) \quad + \quad P(E_1|not\ H) \times P(E_2|not\ H) \times P(not\ H)$$

Hence:

$$P(H|(E_1\ and\ E_2)) = \frac{P((E_1\ and\ E_2)|H) \times P(H)}{P(E_1|H) \times P(E_2|H) \times P(H) \quad + \quad P(E_1|not\ H) \times P(E_2|not\ H) \times P(not\ H)}$$

$$= \frac{0.99 \times 0.99 \times 0.001}{0.99 \times 0.99 \times 0.001 + 0.05 \times 0.05 \times 0.999} = 0.28183 = 28.183\%$$

Prosecutors fallacy

innocent H DNA comes from someone other than suspect

E DNA matches suspects DNA

$$p(E \mid H) = 1/\text{million}$$

As we have evidence E

"There is only 1/million chance"

DNA comes from someone else"

$$p(E \mid H) = p(H \mid E) \qquad p(H) = \frac{9{,}999{,}999}{10{,}000{,}000}$$

For mentimeter

Common enzyme deficiency affects 1 in 200 people

Screening test has:
60% true positive rate
70% true negative rate

Without using any formula or tools answer the following

What is probability you have the enzyme deficiency if you test positive?
What is probability you have the enzyme deficiency if you test negative?

For mentimeter

Calculating probability of suffering dementia

It's claimed footballers are more likely to suffer dementia than ordinary people because of heading the ball throughout their career.

You are responsible for pricing an insurance policy for professional footballers that properly quantifies their risk of getting dementia.

The information available to you is

a) An extensive study of 1000 footballers and random 10,000 non-footballers born between 1965 and 1979 that shows no significant difference in dementia rate between the two groups
b) 10 of the 22 England 1966 world-cup winning squad suffered dementia;
c) a study that claims 15% of adults who live to be 75 suffer dementia;
d) the judgment of a top neurologist who claims heading a ball cannot cause dementia