stat 88 lec 4

warm up 2:00–2:10

a) If you deal 2 cards, what is the chance the 2^{nd} card is red?

i.e find $P(R_2)$

$$P(R_2) = P(R_1R_2) + P(B_1R_2) = \frac{26}{52}\left(\frac{25}{51} + \frac{26}{51}\right)$$

$\frac{51}{51}$

$$P(R_1R_2) = P(R_1)P(R_2|R_1) = \frac{26}{52}\cdot\frac{25}{51}$$

$$P(B_1R_2) = P(B_1)P(R_2|B_1) = \frac{26}{52}\cdot\frac{26}{51}$$

b) Find $P(R_{20}R_{33}) = P(R_{20})P(R_{33}|R_{20}) = \frac{26}{52}\cdot\frac{25}{51}$

33^{rd} card is black

$P(R_{20}B_{33}) = \frac{26}{52}\cdot\frac{26}{51}$

$P(B_{52}|R_{21}R_{40}) = \frac{26}{50}$

Q1 next Thursday (chap 1 and 2)

Last time

Sec 2.1

Inclusion exclusion (OR) rule

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Multiplication (AND) rule

$$P(AB) = P(A)P(B|A)$$

Sec 2.2

When randomly sampling from a population (with or without replacement) be aware of the difference between unconditional and conditional probability.

ex Consider a deck of cards

$$P(B_2) = 26/52$$

$$P(B_2|R_1) = 26/51$$

Today

0. review concept test last time.
1. Sec 2.2 Bayes's rule
2. sec 2.3 Use and interpretation of Bayes's rule.

1 Concept test,

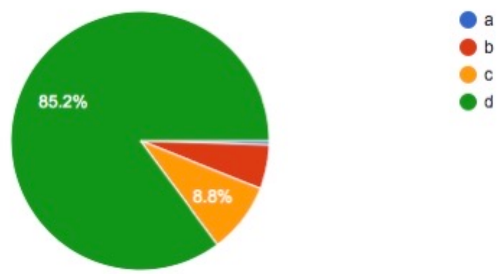
A deck of cards is shuffled. What is the chance that the top card is the **king** of spades **or** the bottom card is the **king** of spades

a $\frac{1}{52}+\frac{1}{52}-\frac{1}{52}\times\frac{1}{52}$

b $\frac{1}{52}+\frac{1}{51}$

c $\frac{1}{52}+\frac{1}{52}-\frac{1}{52}\times\frac{1}{51}$

d none of the above



c	chance of getting a king of spades is 1/52 both at the bottom and top of the deck. subtracting 1/52 * 1/51 is necessary because otherwise the intersection is double counted

d	Since you cannot have both the top and bottom card be king of spades, P(AB) is going to be zero. The correct answer would just be the P(A) + P(B) which would be (1/52 + 1/52)

Sec 2.3 Bayes Rule

Sometimes we can calculate $P(A|B)$ directly

$$\text{ex } P(R_2|B_1) = \frac{26}{51}$$

Here we have 2 stages and we condition on what happens in the first stage.

In backwards conditioning we need the division rule

$$P(AB) = P(A|B)P(B) \Rightarrow P(A|B) = \frac{P(AB)}{P(B)} \quad (\text{assume } P(B) > 0)$$

ex (random container)

A jar contains 3 red balls and 2 green balls

A box contains one red and 4 green balls

Randomly pick a container and a ball from the container.

Find the chance you pick the box if the ball is red

let J = jar

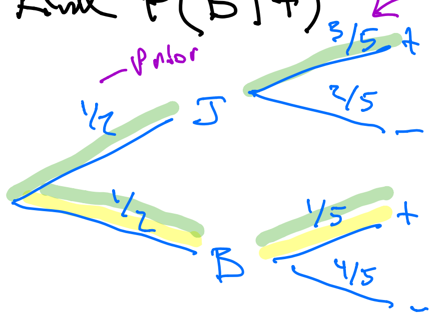
B = box

$+$ = positive test for red

$-$ = negative test for red

likelihood

Find $P(B|+)$ $\leftarrow P(+|J)$



$$P(B|+) = \frac{P(B+)}{P(+)} = \frac{\frac{1}{2}\cdot\frac{1}{5}}{\frac{1}{2}\cdot\frac{1}{5} + \frac{1}{2}\cdot\frac{3}{5}} = \frac{1}{4}$$

posterior prior "updated given data"

We have updated our opinion about whether you picked the box or jar. Before knowing the color of the ball we said chance of drawing the box is 1/2.
After picking the ball and seeing it is red we update the chance of picking the box to 1/4.

about updating chances

Bayes' rule :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Prior knowledge about B, P(A) is called prior probability
given knowledge about B, P(A|B) is called posterior probability

ex Chap 2 (Exercise 2.6.9)

9. A factory has two widget-producing machines. Machine I produces 80% of the factory's widgets and Machine II produces the rest. Of the widgets produced by Machine I, 95% are of acceptable quality. Machine II is less reliable – only 85% of its widgets are acceptable.

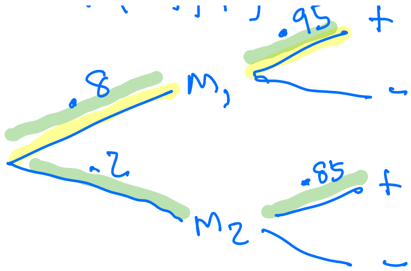
MI = machine I
\+ = positive for acceptable

Suppose you pick a widget at random from those produced at the factory.

a) Find the chance that the widget is acceptable, given that it is produced by Machine I. $P(+|MI) = .95$

b) Find the chance that the widget is produced by Machine I, given that it is acceptable.

$P(MI|+)$ (.8)(.95)



$$P(M_1 \mid +) = \frac{\quad}{(.8)(.95) + (.2)(.85)} = $$

② Sec 2.4 Uses and interpretation of Baye's rule

Harvard Medical School Survey (60 participants)

base rate or prior of disease $= P(D) = 1/1000$

$P(+ \mid \sim D) = .05$

"If a test to detect a disease whose prevalence is 1/1,000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?"

$P(D \mid +)$

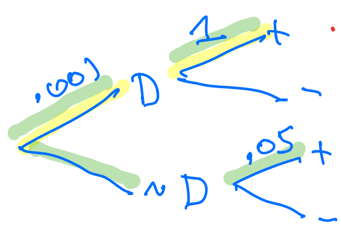
(assume true pos rate is 1) — $P(+ \mid D) = 1$

Harvard med student answers ranged from 2% to 95%

↑ 27/60 said this.

What do you say?

D = disease
$\sim D$ = not disease
$+$ = pos test
$-$ = neg test



$$P(D \mid +) = \frac{(.001)(1)}{(.001)(1) + (.999)(.05)} \approx .02$$

Is this surprising? Yes and no.

Yes if $.02 = 1/50$ seems small to you and you were expecting a bigger number since the true pos rate is 1.

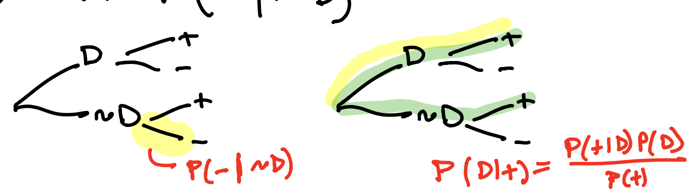
No since $\frac{1}{50}$ is a lot bigger than the prior $\frac{1}{1000}$ and the chance of a false positive is $.05$ which is pretty big.

Base Rate Fallacy: Dont ignore the base rate when computing backward conditionals.

How did so many people get 95%?

$P(-\mid\sim D) = 95\%$

People confuse $P(D\mid+)$ for $P(-\mid\sim D)$



To compute $P(D\mid+)$ you need to take into account the base rate and people tend not to do this, instead focusing on likelihoods such as $P(-\mid\sim D)$.

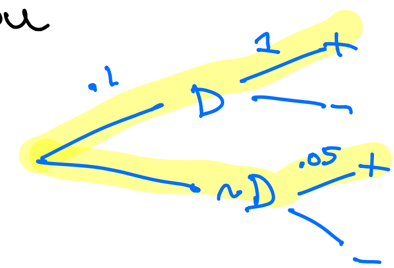
__Key point:__ Posterior prob is effected by base rate as well as the likelihoods.

Particular circumstances effect the prior (base rate).

Suppose you have a 10% chance of having the disease because you show some symptoms or have a family history.

This changes prior to .1 from .001.

Now



$$P(D|+) = \frac{.1(1)}{.1(1) + (.9)(.05)} = .69$$

ex

A True/False test consists of 60 questions. A student knows the answers to 45 of the questions. The remaining 15 answers he guesses at random by tossing a fair coin each time. If it lands heads he answers True and if it lands tails he answers False.

A question is picked at random from the 60 questions on the test. Given that the student got the right answer, what is the chance that he knew the answer?

K = knows

G = guessed

+ = positive for right answer

$$P(K \mid +) = \frac{\frac{3}{4} \cdot 1}{\frac{3}{4} \cdot 1 + \frac{1}{4} \cdot \frac{1}{2}} = \frac{6}{7}$$