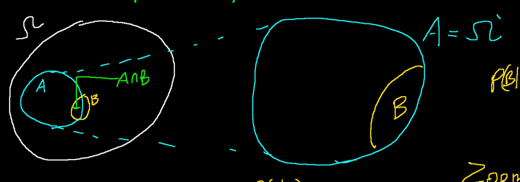
Midterm II

Imagine n = 1000 people (assume they are the universe)
There are 200 smokers (A: smoking)
There are 60 lung cancer diagnoses (B: lung cancer)
36 smokers who have lung cancer (A ∩ B)

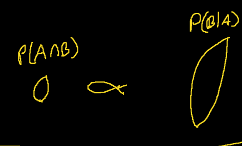
Via the probability definition of long run frequency,
P(A) = 200 / 1000 = 0.200
P(B) = 60 / 1000 = 0.060
P(A ∩ B) = 36 / 1000 = 0.036

What if I want to know the probability of lung cancer only among smokers (given smoking, what is the probability of lung cancer). We denote this P(lung cancer | smoking) = P(B | A). This is called a "conditional probability" because we're conditioning on an event.



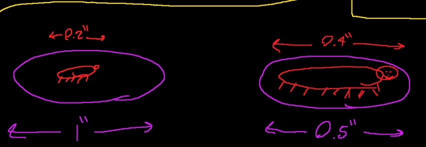
$P(B|A) = \frac{36}{200} = 0.18$

$\text{Zoom} = \frac{P(\Omega)}{P(A)} = \frac{1}{P(A)}$



$\Rightarrow P(B|A) = \text{Zoom}\, P(A,B) = \frac{1}{P(A)} P(A,B)$

$\boxed{P(B|A) = \frac{P(A,B)}{P(A)}}$ definition of conditional prob.



$\text{Zoom} = \frac{\text{prior scope size}}{\text{new scope size}} = \frac{1''}{0.5''} = 2$

$P(A,B) = P(A)P(B|A), \quad P(A|B) = \frac{P(A,B)}{P(B)}$

$\Downarrow$

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ Bayes' Rule (1763)

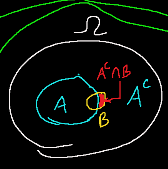
$P(B|A) = \frac{.036}{.2} = .18$

P(smoking | lung cancer) = P(A | B) = .036 / .06 = .6

P(lung cancer | did not smoke) = $P(B \mid A^C) = \frac{P(A^c, B)}{P(A^c)} = \frac{.024}{0.8} = .03$

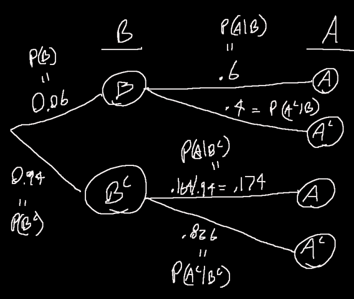
$P(A^c) = 1 - P(A) = 1 - 0.2 = 0.8$

$P(A^c, B) = P(B) - P(A,B)$
$= 0.06 - 0.036 = 0.024$



$\frac{P(B|A)}{P(B|A^c)} = \frac{.18}{.03} = 6$ you are six times more likely to get lung cancer if you smoke than if you don't smoke ("risk ratio")

How many conditional probability questions can be asked about the two events A and B? There are 8. Let's draw two tree illustrations to display all 8 conditional probabilities, the 4 un-conditional probabilities and the 4 joint events.

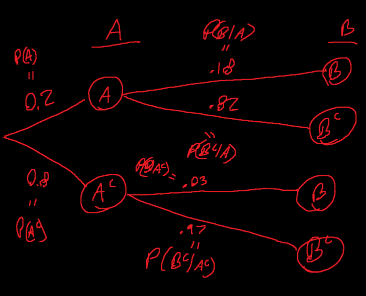


joint events
$P(A,B) = .036$
$P(A^c,B) = .024$
$P(A,B^c) = .164$
$P(A^c,B^c) = .776$

Complement rule for conditional probability: $P(A|B) = 1 - P(A^c|B)$

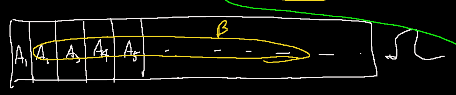
$P(A,B^c) = P(A) - P(A,B) = .2 - .036 = .164$
$P(A^c,B^c) = P(B^c) - P(A,B^c) = .94 - .164 = .776$

To compute the other four conditional probabilities, we need to "invert the tree" considering the event A "first" and B "second".



joint
$P(A,B) = .036$
$P(A,B^c) = .164$
$P(A^c,B) = .024$
$P(A^c,B^c) = .776$

Consider an event B and events A_1, A_2, ... where the A's are mutually exclusive and collectively exhaustive.

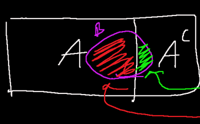


$P(B) = P(B \cap \Omega)$
$= P(B \cap (A_1 \cup A_2 \cup A_3 \cup \ldots\ldots))$
$= P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup \ldots\ldots)$

Note: B ∩ A_1 is disjoint from B ∩ A_2

$(B \cap A_1) \cap (B \cap A_2) = B \cap A_1 \cap B \cap A_2 = B \cap (A_1 \cap A_2) = B \cap \phi = \phi$

$\underset{[A3]}{=} P(B \cap A_1) + P(B \cap A_2) + \ldots\ldots$ Law of Total Probability



$P(B) = P(B,A) + P(B,A^c)$

Conditional probability is very strange and unintuitive. For example: You have two kids and I know at least one is a girl. What is the probability the other is a girl? Intuitively it's 50%! But that's wrong.

$P(\{GG\} \mid$ at least one is a girl$)$
$= P(\{GG\} \mid \{BG, GB, GG\}) = \frac{|\{GG\}|}{|\{BG,GB,GG\}|} = \frac{1}{3}$ all outcomes are equally likely

Thm 5: classic defintion of probability

$|\Omega| < \infty$, if $P(\{\omega_i\}) = \frac{1}{|\Omega|} \Rightarrow P(A) = \frac{|A|}{|\Omega|}$

let $n = |A| < \infty \Rightarrow A = \{\omega_1, \omega_2, \ldots, \omega_n\}$
$= \{\omega_1\} \cup \{\omega_2\} \cup \ldots \{\omega_n\}$

these events are disjoint

$\underset{[A3]}{\Rightarrow} P(A) = P(\{\omega_1\}) + P(\{\omega_2\}) + \ldots + P(\{\omega_n\})$
$= \frac{1}{|\Omega|} + \frac{1}{|\Omega|} + \ldots + \frac{1}{|\Omega|}$
$= \frac{n}{|\Omega|} = \frac{|A|}{|\Omega|}$