Lec 7

$P(A|B) = \dfrac{P(A\cap B)}{P(B)}$ $\quad P(B|A) = \dfrac{P(A\cap B)}{P(A)}$

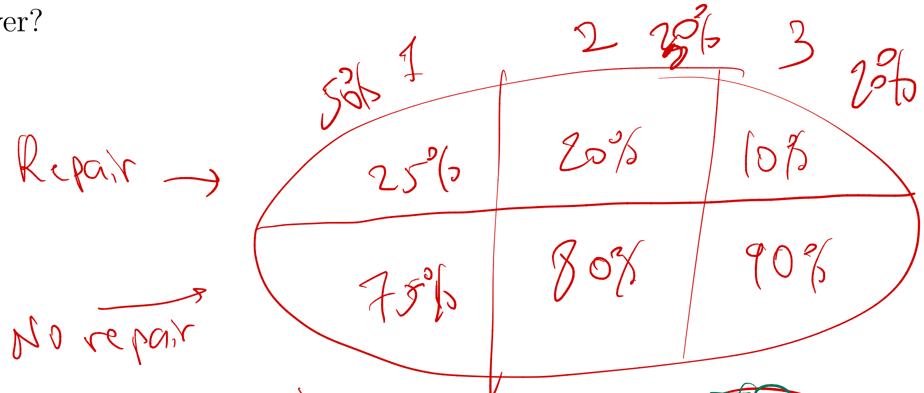
The definition of conditional probability yields the following result.

> The Multiplication rule $P(A\cap B) = P(B)P(A|B)$ and $= P(A)P(B|A)$

Example 14. (Example 2.29 from textbook page 78) An electronics store sells three different brands of DVD players. Of its DVD player sales, 50% are brand 1 (the least expensive), 30% are brand 2, and 20% are brand 3. Each manufacturer offers a 1-year warranty on parts and labor. It is known that 25% of brand 1's DVD players require warranty repair work, whereas the corresponding percentages for brands 2 and 3 are 20% and 10%, respectively.

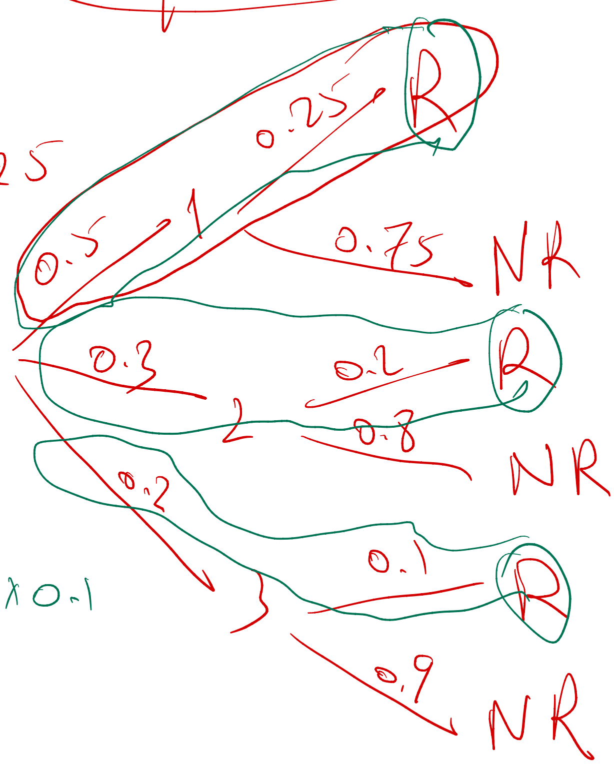
1. What is the probability that a randomly selected purchaser has bought a brand 1 DVD player that will need repair while under warranty?
2. What is the probability that a randomly selected purchaser has a DVD player that will need repair while under warranty?
3. If a customer returns to the store with a DVD player that needs warranty repair work, what is the probability that it is a brand 1 DVD player? A brand 2 DVD player? A brand 3 DVD player?

Solution.



$P(1\cap R) = P(1)P(R|1)$

$= 0.5\times 0.25$



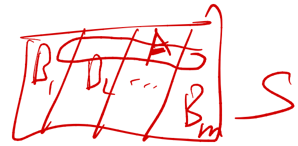
$P(R)$

$= P(1\&R) + P(2\&R) + P(3\&R)$

$= 0.5\times 0.25 + 0.3\times 0.2 + 0.2\times 0.1$

$P(1|R) = \dfrac{P(1\&R)}{P(R)}$

nonoverlapping exhaustive events →



Problem 2 in Example 14 is an example of the Law of Total Probability.

The Law of Total Probability

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_m)P(A|B_m)$$

NOTE: A set of events is jointly **exhaustive** if at least one of the events must occur.

Example 15. (Example 2.30 from textbook page 81) An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account # 1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively.

Question: What is the probability that a randomly selected message is spam?

Solution.

Do it

$P(B_1), P(B_2), \cdots$ & $P(A|B_1), P(A|B_2), \cdots$

Bayes' theorem
Let $\cancel{A_1}B_1, \cancel{A_2}B_2, \cdots, \cancel{A_k}B_m$ be a collection of $\cancel{k}m$ nonoverlapping exhaustive events

____ Then for any other

event $\cancel{B}A$ for which $P(\cancel{B}A) > 0$, the posterior probability of $\cancel{A_j}B_j$ given that $\cancel{B}A$ has occurred is

$$P(B_1|A) = \frac{P(A \cap B_1)}{P(A)} = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots}$$

The transition from the second to the third expression rests on using the multiplication rule in the numerator and the law of total probability in the denominator.

Example 16. (Example 2.31 from textbook page 81) Incidence of a rare disease. Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time (the sensitivity of this test is 99% and the specificity is 98%; in contrast, the Sept. 22, 2012 issue of The Lancet reports that the first at-home HIV test has a sensitivity of only 92% and a specificity of 99.98%).

Question:

If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

Solution.

Do it (example next is similar)

Suppose $P(\text{sick}) = 1\%$

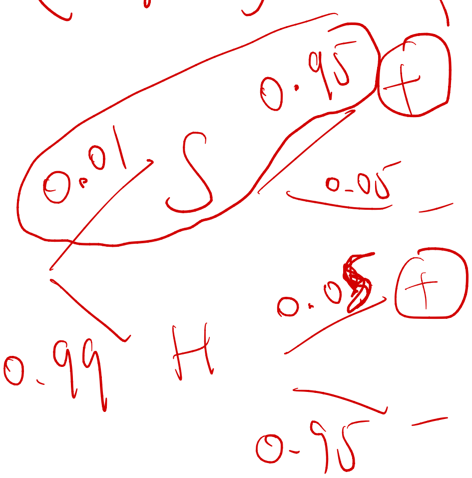
There is a test for the disease.

$P(\text{positive} \mid \text{sick}) = 95\%$ (sensitivity)

$P(\text{negative} \mid \text{not sick}) = 95\%$ (specificity)

$$P(S \mid +) = \frac{P(S \& +)}{P(+)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99}$$

(S = sick, H = healthy)



0.161

16% is low!

So test not accurate enough and disease rare

e.g. false positive 1% instead of 5% (specificity = 99%) gives $P(S \mid +) = 48.97\%$

OR: target high risk group when testing: suppose in this group $P(S) = 40\%$

then $P(S \mid +) = 92.68\%$ w/ original test

PCR vs. antigen?

Antigen: $P(+|S) = 0.8$ (depends on day, brand, etc)

$P(+|S^c) = 0.004$

PCR: $P(+|S) = 0.97$

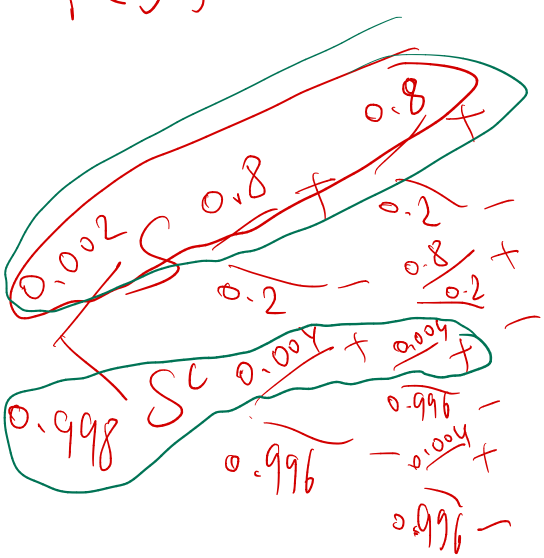
$P(+|S^c) = 0.004$

$P(S) = 0.01 \Rightarrow$ antigen: $P(S|+) = 0.6689$

PCR: $P(S|+) = 0.7101$

$P(S) = 0.5 \Rightarrow$ antigen: $P(S|+) = 0.9950$

PCR: $P(S|+) = 0.9959$



$P(+) = 0.002$ (at peak)

$P(S|+) = 0.2861$

$P(S|++) = 0.9877$

$$\frac{P(S++)}{P(++)} = \frac{P(S++)}{P(S++) + P(S^c++)}$$

$$= \frac{0.002 \times 0.8 \times 0.8}{0.002 \times 0.8 \times 0.8 + 0.998 \times 0.004 \times 0.004}$$