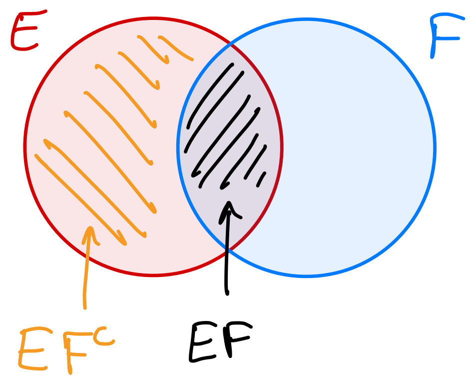
MAT 330/681 Lecture 6 2/22/2021

Bayes Formula



$$E = \underline{EF} \cup \underline{EF^c}$$

$$P(E) = P(EF) + P(EF^c)$$

$$= P(E|F)\cdot P(F) + P(E|F^c)\cdot P(F^c)$$

$$\boxed{P(E) = P(E|F)\cdot \underline{P(F)} + P(E|F^c)\cdot \underline{P(F^c)}}$$

weights in a weighted average of $P(E|F)$ and $P(E|F^c)$

From last time:

$$P(E|F) = \frac{P(EF)}{P(F)} \longrightarrow P(EF) = P(E|F)\cdot P(F)$$

Ex: An insurance company designates people as "accident-prone" or not. Someone that is accident-prone has 40% chance of having an accident in the 1^{st} year of a policy, while someone who is not accident-prone has only half that chance.

Q1: If 30% of the population is accident-prone, what is the chance of a new policy holder having an accident in their first year?

A_1 = having an accident in 1^{st} year

A = being accident-prone.

$P(A) = 0.3$, $P(A^c) = 0.7$

$P(A_1|A) = 0.4$ $P(A_1|A^c) = 0.2$

$$P(A_1) = \underbrace{P(A_1|A)}_{0.4}\underbrace{P(A)}_{0.3} + \underbrace{P(A_1|A^c)}_{0.2}\cdot\underbrace{P(A^c)}_{0.7} = 0.12 + 0.14 = 0.26$$

$\boxed{26\%}$

Q2: If a new policy holder had an accident in their 1st year, what is the prob. they are accident-prone?

$$P(A|A_1) = \frac{P(AA_1)}{P(A_1)} = \frac{P(A_1|A)\cdot P(A)}{P(A_1)} = \frac{0.4 \cdot 0.3}{0.26} = \frac{6}{13}$$

$$P(A_1|A) = \frac{P(A_1 A)}{P(A)} \Rightarrow P(A_1 A) = P(A_1|A)\cdot P(A)$$

$\boxed{46.15\%}$

Ex: Suppose that a multiple choice question in a Final Exam has 5 alternatives, and only 1 is correct. The probability that a student knows the answer to that question is p, and if the student doesn't know it, they guess the answer at random.

a) What is the prob. that the student knew the answer given that they got the correct answer?

K = Knowing the answer
C = getting the correct answer.

$$P(K|C) = \frac{P(KC)}{P(C)} = \frac{\overbrace{P(C|K)}^{1}\cdot\overbrace{P(K)}^{p}}{\underbrace{P(C|K)}_{1}\underbrace{P(K)}_{p} + \underbrace{P(C|K^c)}_{1/5}\cdot\underbrace{P(K^c)}_{(1-p)}}$$

$$= \frac{p}{p + \frac{1}{5}(1-p)} = \frac{5p}{5p+1-p} = \boxed{\frac{5p}{4p+1}}$$

b) What is the prob. that the student got the correct answer and did not know the correct answer?

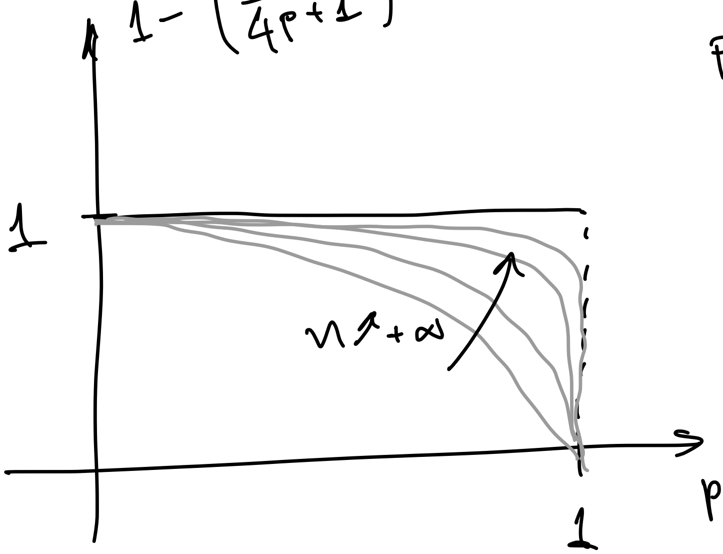
$$P(CK^c) = \underbrace{P(C|K^c)}_{1/5}\underbrace{P(K^c)}_{1-p} = \boxed{\frac{1-p}{5}}.$$

E.g., if $p = \frac{1}{2}$:

a) $\frac{5p}{4p+1} = \frac{5/2}{3} = \boxed{\frac{5}{6}} = \underline{83.33\%}$

b) $\frac{1-p}{5} = \boxed{\frac{1}{10}} = \underline{10\%}$

<u>Curiosity</u>: $1 - \left(\frac{5p}{4p+1}\right)^n$ ← number of questions.

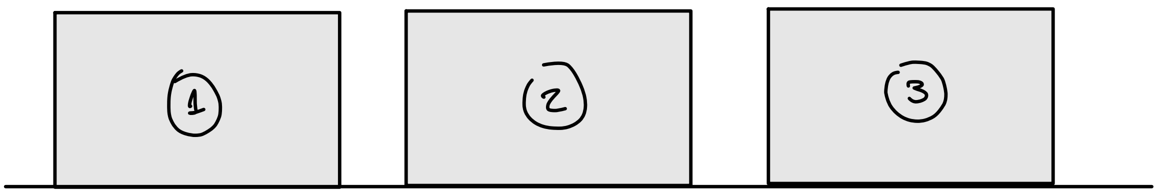


$$P(K^c|C) = 1 - P(K|C) = 1 - \frac{5p}{4p+1}$$

= "prob. you had to guess, given that you got the correct answer".

<u>Upshot</u>: Unless $p = 1$ (you knew everything), as $n \nearrow +\infty$, the prob. you had to guess at least 1 answer given that you got a perfect score goes to 100%..

Monty Hall problem



$H = \left(\begin{array}{c}\text{car is behind}\\ \text{door \#1}\end{array}\right)$ "hypothesis" $P(H) = \frac{1}{3}$, $P(H^c) = \frac{2}{3}$

$E = \left(\begin{array}{c}\text{host opens a door}\\ \text{w/ a goat and no car}\end{array}\right)$ "evidence"

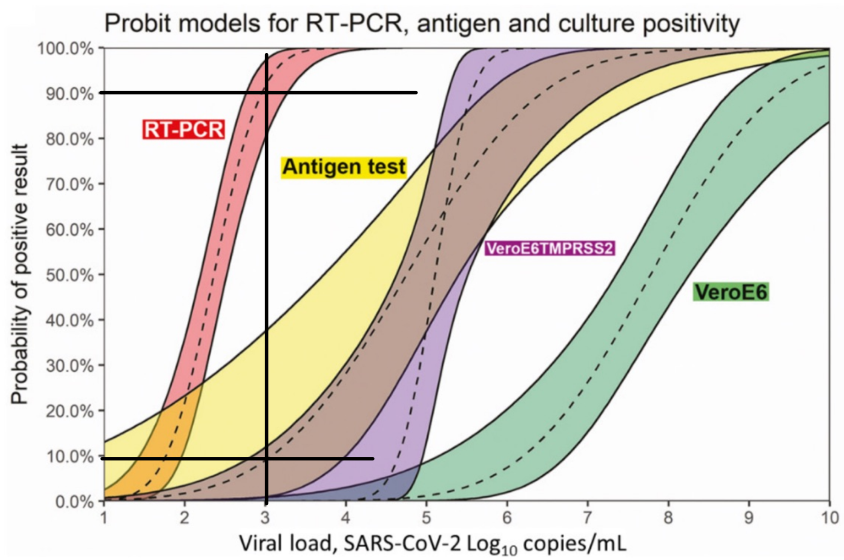
$$\underbrace{P(H|E)}_{\text{prob. of success if we don't switch}} = \frac{P(HE)}{P(E)} = \frac{\overbrace{P(E|H)}^{1} \cdot \overset{=\frac{1}{3}}{P(H)}}{\underbrace{P(E|H)}_{1} \underset{=\frac{1}{3}}{P(H)} + \underbrace{P(E|H^c)}_{1} \cdot \underset{=\frac{2}{3}}{P(H^c)}} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{2}{3}} = \boxed{\frac{1}{3}}$$

$$P(H^c|E) = 1 - P(H|E) = 1 - \frac{1}{3} = \boxed{\frac{2}{3}}$$ ← prob of success if we switch.

Ex: COVID tests

February 5, 2021 Edition 2021-02-05 (75)

Figure.



Note: Adapted from Pekosz *et al.* Modelled probability distributions of each SARS-CoV-2 test type, **RT-PCR**, **antigen**, and viral cultures (**VeroE6TMPRSS2** and **VeroE6**), across log-transformed viral load. As viral load increases, the probability of any test producing a positive result increases.

(With viral load 3:)

- PCR test is positive w/ 90% prob.
- Antigen test is positive w/ 10% prob.
- (False positives occur with 2% prob.)

From Google: (NY Times) Population of NY State: 19.45 million

7-day average of cases per day in NY State: 7,400

$\Rightarrow$ Prevalence of disease: (disease lasts 10 days) $\dfrac{74{,}000}{19{,}450{,}000} \cong 0.004$ (= 0.4%)

Q: If you get a positive PCR test, what is the prob. that you actually have COVID?

D = Have the disease (COVID)

E = Test is positive ("evidence")

PCR:

$$P(D|E) = \frac{P(DE)}{P(E)} = \frac{\overbrace{P(E|D)}^{0.9}\cdot\overbrace{P(D)}^{0.004}}{\underbrace{P(E|D)}_{0.9}\,\underbrace{P(D)}_{0.004} + \underbrace{P(E|D^c)}_{0.02}\cdot\underbrace{P(D^c)}_{0.996}} = \boxed{15.30\%}$$

Antigen:

$$P(D|E) = \frac{\overbrace{P(E|D)}^{0.1} \cdot \overbrace{P(D)}^{0.004}}{\underbrace{P(E|D)}_{0.1} \cdot \underbrace{P(D)}_{0.004} + \underbrace{P(E|D^c)}_{0.02} \cdot \underbrace{P(D^c)}_{0.996}} = \boxed{1.97\%}$$

⚠

How to make better tests?

$$P(D|E) = \frac{P(E|D) \cdot P(D)}{P(E|D) \cdot P(D) + \underbrace{P(E|D^c) \cdot P(D^c)}_{P(ED^c)}}$$

prob. of test being positive given that disease is present.

probability of a false positive

One of the issues with the numbers above is the large prob. of false positives

If we replace:

$$P(E|D^c) = 0.001$$

$$P(D|E) = \frac{0.9 \cdot 0.004}{0.9 \cdot 0.004 + \underbrace{0.001 \cdot 0.996}} = \boxed{78.33\%}$$

making this smaller increases the value of $P(D|E)$