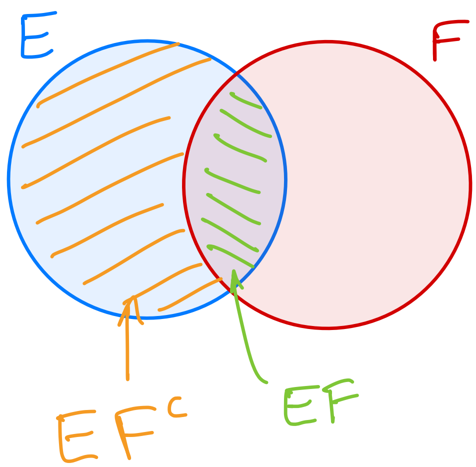
MAT 330/681 Lecture 6 2/16/2022

Bayes' Formula

Recall from last class: $P(E|F) = \dfrac{P(EF)}{P(F)}$ $\quad (P(F) > 0)$



Note: $P(EF) = P(E|F) \cdot P(F)$

$E = EF^c \cup EF$ disjoint union.

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

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

Bayes' Formula:

$$\boxed{P(H|E) = \frac{P(EH)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|H^c) \cdot P(H^c)}}$$

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 1st 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 = being accident-prone

A_1 = having an accident in the 1st yr.

Given: $P(A) = 0.3 \quad P(A^c) = 0.7$

$$P(A_1 | A) = 0.4 \qquad P(A_1^c | A) = 0.6$$

$$P(A_1 | A^c) = 0.2 \qquad P(A_1^c | A^c) = 0.8$$

Want: $P(A_1) = ?$

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

$$\boxed{= 26\%}$$

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

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

$$\boxed{= 46.15\%}$$

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?

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

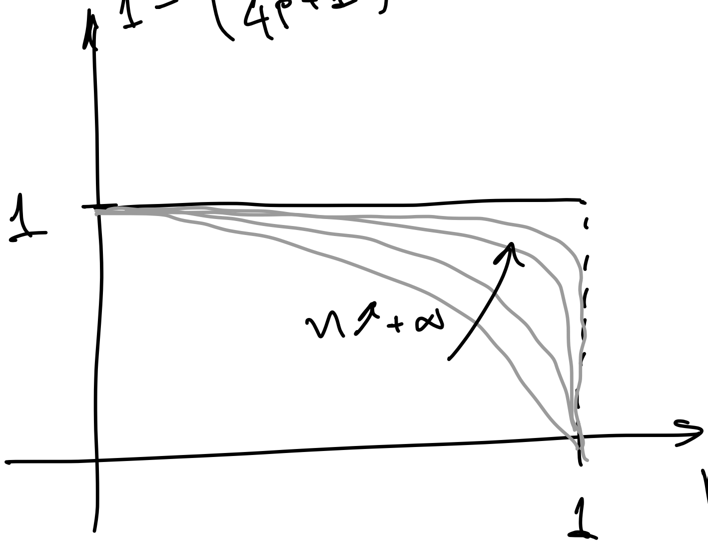
E.g., if $p = \frac{1}{2}$:

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

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

Curiosity:

$$1-\left(\frac{5p}{4p+1}\right)^n \leftarrow \text{number of questions.}$$

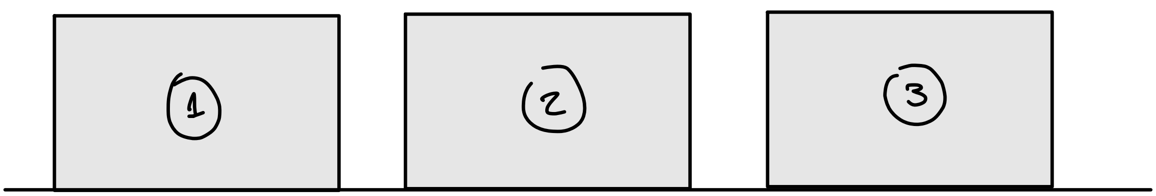


$$P(K^c|C) = 1 - P(K|C) = 1 - \frac{5p}{4p+1}$$

= "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 = $ (car is behind door #1) "hypothesis" $\quad P(H) = \frac{1}{3},\ P(H^c) = \frac{2}{3}$

$E = $ (host opens a door w/ a goat and no car) "evidence"

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

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

False Positives

- The probability of having diabetes is 1% ⇝ $P(D) = 0.01$
- If someone has diabetes, there is a 90% prob. they test positive ⇝ $P(E|D) = 0.9$
- If someone does not have diabetes, the prob they nevertheless test positive is 9% ⇝ $P(E|D^c) = 0.09$ "false positives"
- Someone tests positive. What is the prob. they have diabetes? $P(D|E) = ?$

Most frequent answer among M.D.'s: 80% - 90%

D = disease

E = evidence of the disease (positive test)

$$P(D) = 0.01 \qquad \overset{\text{true positives}}{P(E|D) = 0.9} \qquad \overset{\text{false positives}}{P(E|D^c) = 0.09}$$

$$P(D^c) = 0.99 \qquad \underset{\text{false negatives}}{P(E^c|D) = 0.1} \qquad \underset{\text{true negatives}}{P(E^c|D^c) = 0.91}$$

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

$$= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.09 \cdot 0.99}$$

$$= 9.17\%$$

(cf. 10.11% in Lecture 1)

Q: Why so low?

Tests for a disease are not frequently given to people who don't show any signs/symptoms of that disease. In mathematical terms, this is the difference in prevalence of disease between people who are given tests v. the general population.

Q: How to improve the tests?

- Reduce false positives ($P(E|D^c)$ as small as possible).
- Keep in mind that designing tests for very rare conditions (very small $P(D)$) is much harder.