

DATA 363: Checkpoint FIF2

F1-5

(1, 1) (1, 2) (1, 3) (1, 4)
(2, 1) (2, 2) (2, 3) (2, 4)
(3, 1) (3, 2) (3, 3) (3, 4)
(4, 1) (4, 2) (4, 3) (4, 4)

$$P\{\text{Sum at least 5}\} = \frac{10}{16} \quad P\{\text{first die 2}\} = \frac{4}{16}$$

$$P\{\text{Sum at least 5} \mid \text{first die 2}\} = \frac{2}{4} = \frac{1}{2}$$

This is because out of the 4 outcomes in which the first die is 2, only (2, 3) & (2, 4) give a Sum ≥ 5

F17

$$\begin{aligned} P(\text{Green} \cap \text{Blue} \cap \text{Green}) &= P(\text{Green} \mid \text{Blue} \cap \text{Green}) \times P(\text{Blue} \mid \text{Green}) \\ &\quad \times P(\text{Green}) \\ &= \left(\frac{g-1}{b+g-2} \right) \left(\frac{b}{b+g-1} \right) \left(\frac{g}{b+g} \right) \end{aligned}$$

$$\begin{aligned} P(\text{Green} \cap \text{Blue} \cap \text{Green} \cap \text{Green}) &= P(\text{Blue} \mid \text{Green} \cap \text{Green}) \times P(\text{Green} \mid \text{Green}) \\ &\quad \times P(\text{Green}) \\ &= \left(\frac{b}{b+g-2} \right) \left(\frac{g-1}{b+g-1} \right) \left(\frac{g}{b+g} \right) \end{aligned}$$

~~P(exactly 2 green)~~

$$P(\text{exactly 2 out of 3 green}) = P(B \cap G \cap G) + \overset{P}{P}(G \cap B \cap G) + P(G \cap G \cap B)$$

$$\begin{aligned} \cancel{P(B \cap G \cap G)} \quad P(G \cap G \cap B) &= P(G|G \cap B) * P(G|B) * P(B) \\ &= \left(\frac{g-1}{b+g-2} \right) \left(\frac{g}{b+g-1} \right) \left(\frac{b}{b+g} \right) \end{aligned}$$

$$P(\text{exactly 2/3 Green}) = \boxed{\frac{3(b)(g)(g-1)}{(b+g)(b+g-1)(b+g-2)}}$$

Note: We could add them since they're mutually exclusive outcomes.

$$\begin{aligned} P(\text{exactly 2/4 Green}) &= \frac{\text{Ways of choosing 2G \& 2B}}{\text{Ways of choosing 4 balls}} \\ &= \frac{C(b, 2) \times C(g, 2)}{C(b+g, 4)} \end{aligned}$$

$$= \boxed{\frac{b(b-1)(g)(g-1)}{(b+g)(b+g-1)(b+g-2)(b+g-3)}}$$

Note: We could also write $P(2/3 \text{ green})$ as $\Rightarrow \frac{C(b, 1) \times C(g, 2)}{C(b+g, 3)}$

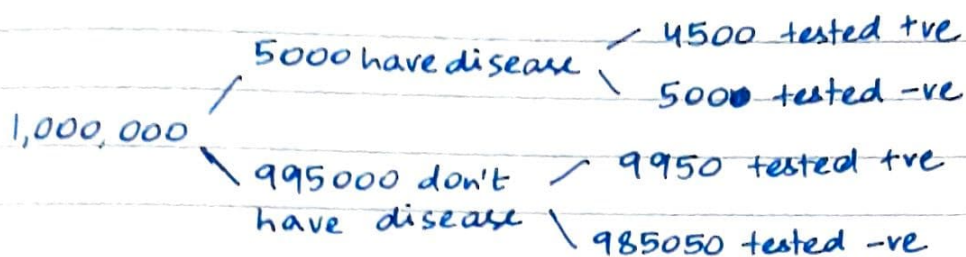
This would give us the same answer.

F2-10

Probability tree for: Percent with disease = ~~1%~~ 0.5%.

False positive = 1%.

True positive = 90%.



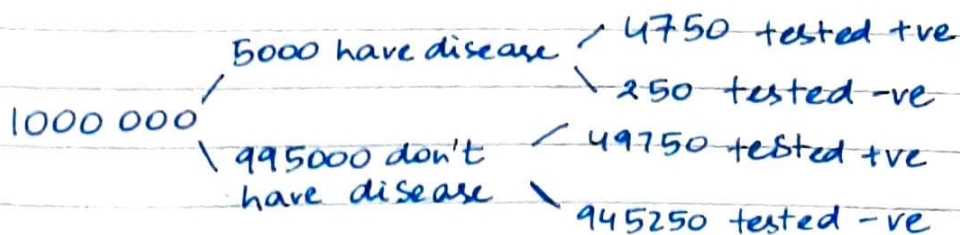
$$P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{P\{\text{people w/disease \& tested +ve}\}}{P\{\text{people tested +ve}\}}$$

$$= \frac{4500 / 1000000}{(9950 + 4500) / 1000000} = \frac{4500}{14450} = \boxed{0.3114}$$

Probability tree for: Percent with disease = 0.5%.

False positive = 5%.

True positive = 95%.



$$P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{4750 / 1000000}{(4750 + 49750) / 1000000} = \frac{4750}{54500} = \boxed{0.0871}$$

A change in the false positive rate has a bigger impact. This makes sense intuitively since the false +ve rate acts on a significantly LARGER number, meaning even a small change in the percentage would lead to a large change in the actual number of people.

The true tre rate on the other hand affects a smaller part of the population. Therefore even a drastic change would not warrant for a large inc or dec in the actual number of people.