DATA 363: Checkpoint F1F2

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

F1-7

$P(\text{Green} \cap \text{Blue} \cap \text{Green}) = P(\text{Green} \mid \text{Blue} \cap \text{Green}) \times P(\text{Blue} \mid \text{Green}) \times P(\text{Green})$

$$= \left(\frac{g-1}{b+g-2}\right)\left(\frac{b}{b+g-1}\right)\left(\frac{g}{b+g}\right)$$

$P(\text{Blue} \cap \text{Green} \cap \text{Green}) = P(\text{Blue} \mid \text{Green} \cap \text{Green}) \times P(\text{Green} \mid \text{Green}) \times P(\text{Green})$

$$= \left(\frac{b}{b+g-2}\right)\left(\frac{g-1}{b+g-1}\right)\left(\frac{g}{b+g}\right)$$

~~P(exactly 2 green)~~

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

~~P(G∩G∩B)~~ $P(G \cap G \cap B) = P(G \mid G \cap B) * P(G \mid B) * P(B)$

$$= \left(\frac{g-1}{b+g-2}\right)\left(\frac{g}{b+g-1}\right)\left(\frac{b}{b+g}\right)$$

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

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

$$P(\text{exactly } 2/4 \text{ Green}) = \frac{\text{Ways of choosing } 2G \,\&\, 2B}{\text{Ways of choosing 4 balls}}$$

$$= \frac{C(b,2) \times C(g,2)}{C(b+g,4)}$$

$$\boxed{= \frac{6(b)(b-1)(g)(g-1)}{(b+g)(b+g-1)(b+g-2)(b+g-3)}}$$

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

This would give us the same answer.

F2-10 Probability tree for: Percent with disease = ~~10%~~ 0.5%
False positive = 1%
True positive = 90%

1,000,000
- 5000 have disease
  - 4500 tested +ve
  - 5000 tested -ve
- 995000 don't have disease
  - 9950 tested +ve
  - 985050 tested -ve

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

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

Probability tree for: Percent with disease = 0.5%
False positive = 5%
True positive = 95%

1000 000
- 5000 have disease
  - 4750 tested +ve
  - 250 tested -ve
- 995000 don't have disease
  - 49750 tested +ve
  - 945250 tested -ve

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

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 tve 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.