Student: Quoc Tuan Vinh, Ngo
ID: 704526

Problem 1. "Coins"

From the question, we have:

- $p(X = t \mid C = c_1) = \theta_1$
- $p(X = t \mid C = c_2) = \theta_2$
- $p(C = c_1) = \pi_1$

Since c_1 & c_2 are random variables $\in \{c_1, c_2\}$, we have $p(C = c_1) + p(C = c_2) = 1$

$\Leftrightarrow p(C = c_2) = 1 - \pi_1$

Since h & t are random variables $\in$ {h, t}, we have:

- $p(X = h \mid C = c_1) + p(X = t \mid C = c_1) = 1 \Leftrightarrow p(X = h \mid C=c_1) = 1 - \theta_1$
- Similarly: $p(X = h \mid C = c_2) = 1 - \theta_2$

Applying Bayes conditional distribution:

$$p(C = c_1 \mid X = t) = \frac{p(X=t \mid C=c1) * p(C=c1)}{p(X=t)} = \frac{\theta_1 \pi_1}{p(X=t)} \quad (1)$$

However, we also have:

$$p(C = c_1 \mid X = t) = 1 - p(C = c_2 \mid X = t) = 1 - \frac{p(X=t \mid C=c2) * p(C=c2)}{p(X=t)} = 1 - \frac{\theta_2 (1-\pi_1)}{p(X=t)} \quad (2)$$

From (1), (2), we can calculate $p(X = t) = \theta_1 \pi_1 + \theta_2 - \pi_1\theta_2$ (3)

Apply (3) → (1): $\mathbf{p(C = c_1 \mid X = t)} = \frac{\boldsymbol{\theta_1 \pi_1}}{\boldsymbol{\theta_1\pi_1 + \theta_2 - \pi_1\theta_2}}$

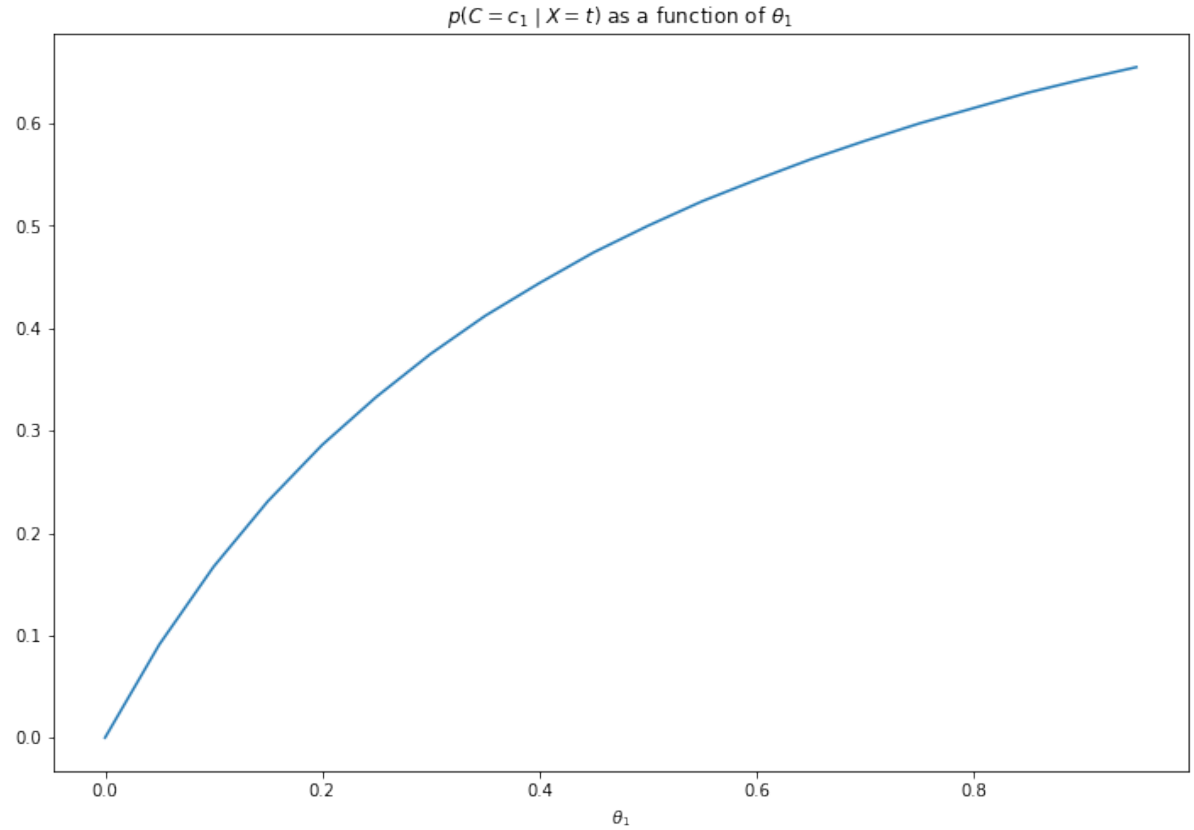
And $\mathbf{p(C = c_1 \mid X = h)} = \frac{p(X=h \mid C=c1) * p(C=c1)}{p(X=h)} = \frac{(1-\theta_1)\pi_1}{1 - p(X=t)} = \frac{\boldsymbol{(1-\theta_1)\pi_1}}{\boldsymbol{1 - \theta_1\pi_1 - \theta_2 + \pi_1\theta_2}}$

Plot:

```
# compute posterior probability of c_1
theta_1 = np.linspace(0,1, num=20, endpoint=False) #(Stepwise of 0.05 between 0 and 0.95)
post_c1 = []
for i in theta_1:
   post_c1.append(np.round((i*pi_1)/(i*pi_1 + theta_2 - pi_1*theta_2),2))
```

Theta_1 values are: [0.   0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65
 0.7 0.75 0.8 0.85 0.9 0.95]

$p(C = c_1 \mid X = t)$ values are: [0.0, 0.091, 0.167, 0.231, 0.286, 0.333, 0.375, 0.412, 0.444, 0.474
, 0.5, 0.524, 0.545, 0.565, 0.583, 0.6, 0.615, 0.63, 0.643, 0.655]



Problem 2. "False positive paradox."

Let e_i , y_i denote the actual and detected outcomes respectively. Thus, y_i has two possibilities of Y (Correctly detected) and N (wrongly detected). Also, e_i also has two possibilities of H (students being honest in reality) and C (cheating students in reality).

From the description, we can easily realize that both of **p(Y | H) = p(N | C) = 0.98** (given the actual outcome, we have the detected probabilities). From there, we can calculate **p (N |H) and p(Y | C) = 0.02** since the variables are random and p(Y | H) + p (N |H) = p(N | C) + P(Y | C) = 1.

Also, from the actuality, we also know that **p(C) (the actual percentage of cheating students) is 1/300 ~ 0.003333**, which means that p(H) is 0.996667.

According to Bayes conditional probability formulas, we know that:

$p(y_i , e_i) = p (y_i \mid e_i) * p (e_i)$, so it is possible to compute $p(y_i , e_i)$ and $p(y_i)$.

From there, it is easy to compute $p(e_i \mid y_i) = p (y_i,e_i) / p(y_i)$. To make it easier, the table below is built:

			p (y_i | e_i)			p (y_i,e_i)		p (e_i|y_i)	
			Detected (y_i)		**Σ**	**Detected (y_i)**		**Detected (y_i)**	
			Y	**N**		**Y**	**N**	**Y**	**N**
Actual (e_i)	**H**	0.996667	0.980000	0.020000	1	0.976733	0.019933	0.999932	0.859195
	C	0.003333	0.020000	0.980000	1	0.000067	0.003267	0.000068	**0.140805**
Σ		1				0.976800	0.023200	1	1

The question asks "If the detector now flags a particular student X as a cheater, how likely is it that X has, in fact, cheated in the exam?", which means that given y_i = N, what is the probability of e_i = C. In this case, **p (C | N) = 0.140805 ~ 14.08%**.

Problem 3. "Markov blanket definition."

The Markov blanket of a node is its parents, children and the parents of its children (excluding itself).

Node A: C, B, D, F

Node B: A, C, D, E, F

Node C: A, B, D

Node D: A,C,B,E,F

Node E: B, D, F

Node F: A, B, D, E.