CA318
Labsheet #6
Solutions

Question 1

The IMF is concerned about organized crime and money transfers. They have analyzed a recent data set as follows, and classified the data as **Criminal** (Y/N). They have the following tests: **Cash** (Y/N), **Currency** (EUR/USD/CAN), **Business Account** (Y/N/?) and **Frequency** (monthly/daily/weekly).

Cash	Currency	Business Acct	Frequency	Criminal?
Yes	EUR	?	monthly	No
Yes	USD	Yes	monthly	No
No	USD	?	monthly	Yes
No	CAN	No	daily	Yes
No	CAN	?	weekly	Yes
No	EUR	Yes	daily	No
No	USD	Yes	daily	No
Yes	CAN	?	weekly	No

Part 1:

Using the **Identity Trees** process from the lectures, rank each of these tests from best to worst according to the number of outcomes placed in homogeneous groups. **Only one iteration is required**. Be sure to show the total number of outcomes placed in homogeneous groups for each test.

1. Business Acct: 4

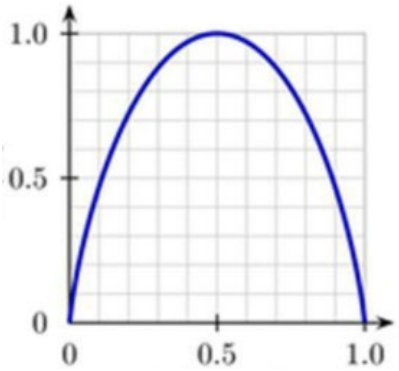
2. Cash: 3
3. Currency: 2
4. Frequency: 0

Part 2:

Now use the disorder distribution computation to the rank quality of each test, showing the total disorder for each test and demonstrating which one is best **and compare your results to the results found in part 1**.

Do your findings here agree with part 1?

Disorder Distribution:



Suggested Approach

You can use the following table structure (**only one iteration required**).

Here is a sample:

Test	Outcome	Weight	P/T	Disorder	Weighted Disorder
Cash	Y	3/8	0/3	0	0
	N	5/8	3/5	0.9	0.562
Currency	USD	3/8	1/3	0.8	0.3
	CAN	3/8	2/3	0.9	0.3375
	EUR	2/8	0/8	0	0
Business Acct	Y	3/8	0/3	0	0
	N	1/8	1/1	0	0
	?	4/8	1/2	1	0.5
Frequency	daily	3/8	1/3	0.8	0.3
	weekly	2/8	1/2	1	0.25
	monthly	3/8	1/3	0.9	0.3375

Ranking, best to worst:

1. Business = 0 + 0 + 0.5 = 0.5
2. Cash = 0 + 0.562 = 0.562
3. Currency = 0 + 0.3 + 0.3375 = 0.6375
4. Frequency = 0.25 + 0.3 + 0.3375 = 0.8875

So yes, the two methods agree where the objective here is to minimize the disorder measure