Artificial Intelligence: Assignment 5, Ege Ersu (ee31) and Linus Shih (yls2)

Question 1

Problem 1

N : number of stars M1, M2 : measurements

F_1 : event that telescope1 undercounts 3+ stars $P(F_1) = f$

F_2 : event that telescope2 under counts 3+ stars $P(F_2) = f$

$P(M_1 = N+1) = e$, $P(M_1 = N-1) = e$, $P(M_1 = N) = 1-2e$

a) X 1 does not. N is the true number of stars and should be a root cause. Also N is not independent of focus F given M_1 and M_2.

✓ 2 represents the intuitive causal structure. True number of stars and the event of an error must effect the measurement. Also distinct telescopes are modeled to be independent.

3 is complicated but still correct.

b) 2 is the best network. It has fewer connections than 3, having less parameters is good. It also resembles the intuitive causal structure.

c) $P(M_1|N) = P(M_1|N, F_1) \cdot P(F_1|N) + P(M_1|N, \overline{F_1}) \cdot P(\overline{F_1}|N)$

$= P(M_1|N, F_1) \cdot P(F_1) + P(M_1|N, \overline{F_1}) \cdot P(\overline{F_1})$

$M_1 = 0$ $N = 1$ $P(M_1 = 0|N=1) = P(M_1 = 0|N=1, F_1) \cdot P(F_1) + P(M_1 = 0|N=1, \overline{F_1}) \cdot P(\overline{F_1})$

$= 1 \cdot f + e \cdot (1-f) = f + e(1-f)$

$M_1 = 0$ $N = 2$ $P(M_1 = 0|N=2) = P(M_1 = 0|N=2, F_1) \cdot P(F_1) + P(M_1 = 0|N=2, \overline{F_1}) \cdot P(\overline{F_1})$

$= 1 \cdot f + 0 \cdot (1-f) = f$

$M_1 = 0$ $N = 3$ $P(M_1 = 0|N=3) = P(M_1 = 0|N=3, F_1) \cdot P(F_1) + P(M_1 = 0|N=3, \overline{F_1}) \cdot P(\overline{F_1})$

$= 1 \cdot f + 0 \cdot (1-f) = f$

$M_1 = 1$ $N = 1$ $P(M_1 = 1|N=1) = P(M_1 = 1|N=1, F_1) \cdot P(F_1) + P(M_1 = 1|N=1, \overline{F_1}) \cdot P(\overline{F_1})$

$= 0 \cdot f + (1-2e) \cdot (1-f) = (1-2e)(1$

$M_2 = 1$ $N = 2$ $P(M_1|N) = P(M_1|N, F_1) \cdot P(F_1) + P(M_1|N, \overline{F_1}) \cdot P(\overline{F_1})$

$= 0 \cdot f + e \cdot (1-f) = e(1-f$

$M_2 = 1$ $N = 3$ $P(M_1|N) = 0 \cdot f + 0 \cdot (1-f) = 0$

$$P(M_1 \mid N) = P(M_1 \mid N, F_1) \cdot P(F_1) + P(M_1 \mid N, \overline{F_1}) \cdot P(\overline{F_1})$$

$M_1 = 2$ $N = 1$	0	$\cdot$ f	+	e	$\cdot$ (1-f)	=	e(1-f)
$M_1 = 2$ $N = 2$	0	$\cdot$ f	+	(1-2e)	$\cdot$ (1-f)	=	(1-2e)(1-f)
$M_1 = 2$ $N = 3$	0	$\cdot$ f	+	e	$\cdot$ (1-f)	=	e(1-f)
$M_1 = 3$ $N = 1$	0	$\cdot$ f	+	0	$\cdot$ (1-f)	=	0
$M_1 = 3$ $N = 2$	0	$\cdot$ f	+	e	$\cdot$ (1-f)	=	e(1-f)
$M_1 = 3$ $N = 3$	0	$\cdot$ f	+	(1-2e)	(1-f)	=	(1-2e)(1-f)
$M_1 = 4$ $N = 1$	0	$\cdot$ f	+	0	$\cdot$ (1-f)	=	0
$M_1 = 4$ $N = 2$	0	$\cdot$ f	+	0	(1-f)	=	0
$M_1 = 4$ $N = 3$	0	$\cdot$ f	+	e	(1-f)	=	e(1-f)

	N = 1	N = 2	N = 3
$M_1 = 0$	1·f + e(1-f)	f	f
$M_1 = 1$	(1-2e)(1-f)	(e)(1-f)	0
$M_1 = 2$	e(1-f)	(1-2e)(1-f)	(e)(1-f)
$M_1 = 3$	0	(e)(1-f)	(1-2e)(1-f)
$M_1 = 4$	0	0	(e)(1-f)

d) Suppose $M_1 = 1$ and $M_2 = 3$

Case 1: $\overline{F_1}, \overline{F_2} \Rightarrow \boxed{N = 2}$ must be the case. With prob e astro1 will count 1 less, with prob e astro will count 1 extra.

Case 2: $\overline{F_1}, F_2 \Rightarrow$ N can't take any value, so this is not possible since F_2 is True, $N \geq M_2 + 3 = 6$. For any $N \geq 6$ M_1 cannot be 1 unless F_1 is True, but it's False

Case 3: $F_1, \overline{F_2} \Rightarrow N \geq M_1 + 3 = 4$. For $N = 4$ if astro 2 observes 1 less with prob e, this can happen. But for any $N > 4$, $M_2 \neq 3$ without F_2 being true. So only $\boxed{N = 4}$

Case 4: $F_1, F_2 \Rightarrow N \geq M_1 + 3 = 4$ and $N \geq M_2 + 3 = 6$.
Any $\boxed{N \geq 6}$ is fine since constraints above are satisfied.

So possible N values are $= \{2, 4, 6, 7, 8, \cdots\}$

e) Most Likely N ?

We need knowledge of prior distribution $P(N = n)$ for all possible n's calculated on part (d).
Let them be $d_2, d_4, d_6, d_7, \dots$
Now posteriors become :

$N = 2$: $d_2 \cdot e^2 (1-f)^2$
$N = 4$: $d_4 \cdot e.f$
$N = 6$: $d_6 \cdot f^2$
$N = 7$: $d_7 \cdot f^2$
$\vdots$

We know $e >> f$, so $N = 2$ seems the most likely due to e^2 and $(1-f)^2$ which are both large.
But info on $d_2, d_4, d_6, d_7 \dots$ are necessary for a proper proof.

Question 2

Gibbs Sampling

2b)

a) Markov Chain will have 4 states

non-evidence variables: Cloudy, Rain

Each can take T/F $\Rightarrow$ $2 \cdot 2 = 4$ states

b) $P(C|R,S) = \alpha \cdot P(C) \cdot P(S|C) \cdot P(R|C)$

$= \alpha \cdot \langle 0.5, 0.5 \rangle \cdot \langle 0.1, 0.5 \rangle \cdot \langle 0.8, 0.2 \rangle$

$= \alpha \langle 0.04, 0.05 \rangle$

$= \langle 4/9, 5/9 \rangle$

$P(C|\bar{R},S) = \alpha \cdot P(c) \cdot P(S|C) \cdot P(\bar{R}|C)$

$= \alpha \langle 0.5, 0.5 \rangle \langle 0.1, 0.5 \rangle \cdot \langle 0.2, 0.8 \rangle$

$= \langle 1/21, 20/21 \rangle$

$P(R|C,S,W) = P(R|C) \cdot P(W|S,R)$

$= \langle 0.8, 0.2 \rangle \langle 0.99, 0.90 \rangle \Rightarrow$ normalize

$= \langle 22/27, 5/27 \rangle$

c) as $n \to \infty$, probability in being in each state will converge to the values we found in part a using variable elimination.

d) $N = 1000$: 0.5535

$N = 5000$: 0.4087

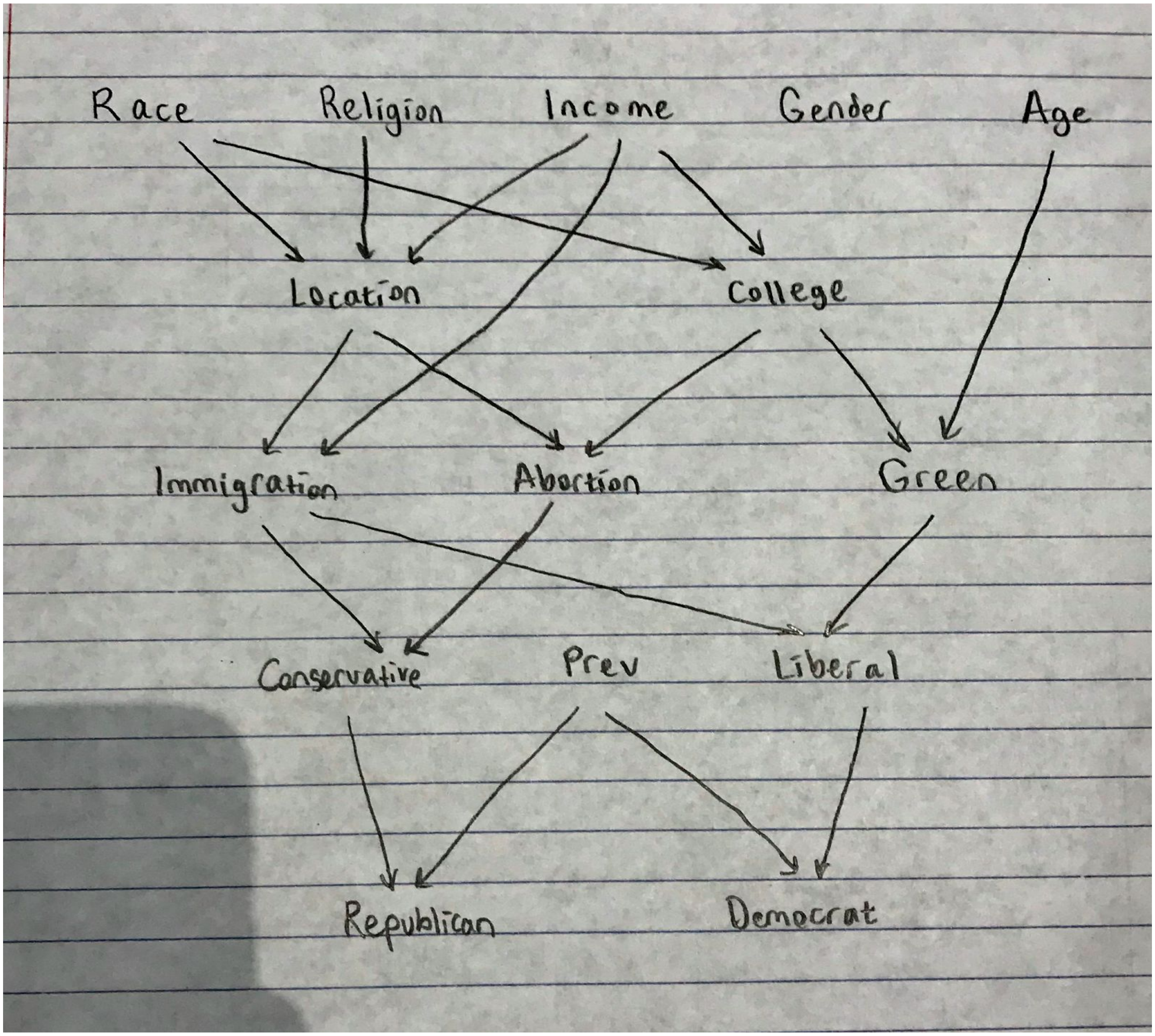
$N = 10000$: 0.4167

for $n = 20{,}000$ looks good enough

Question 3
Particle Filtering
Autograder Results:
----- START PART 3.1a-0
----- END PART 3.1a-0 [took 0:00:00.006078, 10/10 points]
----- START PART 3.2a-0
----- END PART 3.2a-0 [took 0:00:00.010103, 10/10 points]
----- START PART 3.3a-0
----- END PART 3.3a-0 [took 0:00:00.009999, 10/10 points]
----- START PART 3.3a-1
----- END PART 3.3a-1 [took 0:00:00.013025, 10/10 points]
----- START PART 3.3a-2
----- END PART 3.3a-2 [took 0:00:00.016063, 10/10 points]
========== END GRADING [50/50 points]
Total max points: 50

Question 4

The code is in election.py



Test Cases

g.q(race=True, religion=True, income=True, age=True, gender=True)

=> Republican: 69.429482, Democrat: 65.128492

g.q(race=False, religion=True, income=False, age=False, gender=False)

=> Republican: 41.482924, Democrat: 71.284457

g.q(race=True, religion=True, income=True, age=True, gender=False, location=True, college=True)

=> Republican: 21.842942, Democrat: 89.28484

Variables

Race = {True:White, False:Other}, Observable
Religion = {True:Christian, False:Other}, Observable
Income = {True:Middle Class + Rich, False: Working Class + Poor}, Observable
Gender = {True:Male, False:Female}, Observable
Age = {True:Young Voters, False:Other}, Observable
Location = {True:City+Suburb, False:Rural+Country}, Sometimes Observable
College = {True:Went to College, False: Did Not}, Sometimes Observable
Immigration = {True: For, False: Against}, Not Observable
Abortion = {True: For, False: Against}, Not Observable
Green = {True: Care, False: Do not Care}, Not Observable
Conservative = {True,False}, Not Observable
Prev = {True: voted Democrat last election, False: voted Republican last election}, Observable
Liberal = {True, False}, Not Observable
Republican = {True, False}, Not Observable
Democrat = {True, False}, Not Observable

Edges

Location will depend on Race, Religion, Income. We know rural/country areas are mostly white and christian.

College will depend on Race and Income. If the income of a family is higher their kid will go to college with higher chance. If the race is not white, it will be unlikely that the kid will go to college, even more unlikely if they have lower income.

Immigration will depend on Location and Income. Cities are more likely to be for immigration whereas the country is mostly against.

Abortion will depend on Location and College. Educated people are more likely to be for abortion. Cities are more likely to be for abortion whereas rural areas are usually against.

Green will depend on college and age. Educated people have a deeper understanding of climate problems so they are more likely to care. Also younger people care more about the environment than older people who are not inclined to think long-term.

Conservative will depend on immigration and abortion. It’s an oversimplification but immigration reflects how right they lean, and abortion reflects how their daily thoughts are affected by religion.

Liberal will depend on immigration and abortion. Again immigration being an indication of how left they lean, and green being a reflection of their overall mental attitude towards the world. Again a gross oversimplification.

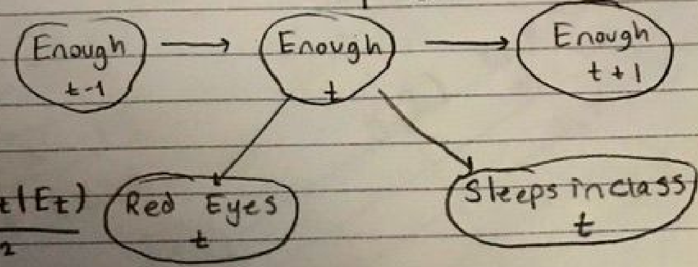
Republican and Democrat nodes depend on Conservative and Liberal nodes, as well as the prev node which reflects what they voted for in the past election. Even if a person has higher prob on the liberal node, if they voted republican in the last election, there is now lower chance that they will vote Democrat, compared to the case in which if we didn’t know what they voted for.

Question 5

E_t	$P(E_t \mid E_{t-1})$
t	0.8
f	0.3

Problem 5

5-1)



E_t	$P(R_t \mid E_t)$
t	0.2
f	0.7

E_t	$P(S_t \mid E_t)$
t	0.1
f	0.3

Let E_t denote enough sleep at time t, the hidden variable.
Let R_t denote having red eyes at time t, an evidence variable
Let S_t denote sleeping in class at time t, an evidence variable

- $P(R_t, S_t \mid E_t) = P(R_t \mid E_t) \cdot P(S_t \mid E_t)$
  $= 0.2 \cdot 0.1 = 0.02$

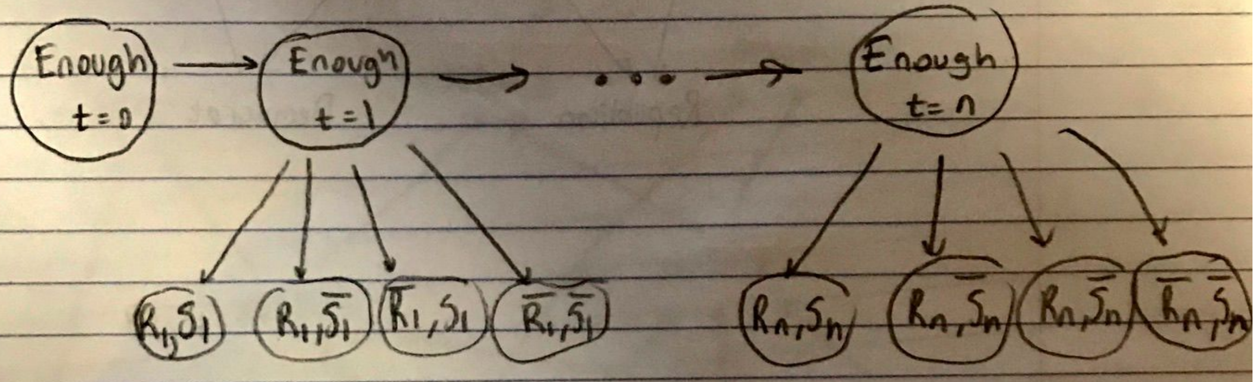
- $P(\overline{R_t}, S_t \mid E_t) = P(\overline{R_t} \mid E_t) \cdot P(S_t \mid E_t)$
  $= 0.8 \cdot 0.1 = 0.08$

- $P(R_t, \overline{S_t} \mid E_t) = P(R_t \mid E_t) \cdot P(\overline{S_t} \mid E_t)$
  $= 0.2 \cdot 0.9 = 0.18$

- $P(\overline{R_t}, \overline{S_t} \mid E_t) = P(\overline{R_t} \mid E_t) \cdot P(\overline{S_t} \mid E_t)$
  $= 0.8 \cdot 0.9 = 0.72$

R_t	S_t	$P(R_t, S_t \mid E_t)$
T	T	0.02
T	F	0.08
F	T	0.18
F	F	0.72

Hidden Markov Model



5-2) e_1 = not red eyes, not sleeping
e_2 = red eyes, not sleeping
e_3 = red eyes, sleeping

$P(E_t \mid e_{1:t})$ for $t = 1,2,3$?

- $P(E_0) = 0.7$, $P(\overline{E_0}) = 0.3$ $\qquad \alpha = 1.85$

- $P(E_1) = \sum_{E_0} P(E_1 \mid E_0) \cdot P(E_0) = 0.8 \cdot 0.7 + 0.3 \cdot 0.3$
$= 0.65$

$P(\overline{E_1}) = 0.35$

$P(E_1 \mid e_1) = \alpha \cdot P(e_1 \mid E_1) \cdot P(E_1)$
$= \alpha \, (0.72)(0.65) = 0.86$

- $P(E_2 \mid e_1) = P(E_2 \mid E_1) P(E_1 \mid e_1) + P(E_2 \mid \overline{E_1}) \cdot P(\overline{E_1} \mid e_1)$
$= 0.8 \cdot 0.86 \qquad 0.3 \cdot 0.14$
$= 0.73$

- $P(E_2 \mid e_{1:2}) = \alpha \, P(e_2 \mid E_2) \cdot P(E_2 \mid e_1)$
$= \alpha \cdot 0.37 \qquad 0.73$
$= 0.50$

- $P(E_3 \mid e_{1:2}) = P(E_3 \mid E_2) \, P(E_2 \mid e_{1:2}) + P(\overline{E_3} \mid E_2) \cdot P(E_2 \mid e_{1:2})$
$= 0.55$

- $P(E_3 \mid e_{1:3}) = \alpha \cdot P(e_3 \mid E_3) \cdot P(E_3 \mid e_{1:2})$
$= 0.10$

5-3) $P(E_t | e_{1:3}) = ?$

- $P(e_3 | E_3) = (0.2 \cdot 0.1, 0.7 \cdot 0.3) = (0.02, 0.21)$
- $P(e_3 | E_2) = (0.02 \cdot 0.8 + 0.21 \cdot 0.2, 0.02 \cdot 0.3 + 0.21 \cdot 0.7)$
  $= (0.059, 0.153)$
- $P(e_{2:3} | E_1) = (0.023, 0.056)$

$\Rightarrow$ namalize $\Rightarrow$

- $P(E_1 | e_{1:3}) = \alpha \cdot P(E_1 | e_1) \cdot P(e_{2:3} | E_1)$
  $= (0.73, 0.27)$
- $P(E_2 | e_{1:3}) = \alpha \cdot P(E_2 | e_{1:2}) \cdot P(e_3 | E_1)$
  $= (0.28, 0.72)$
- $P(E_3 | e_{1:3}) = (0.10, 0.90)$

5-4) for $t=1$, filter prob is 0.86 $\}$ >
smoothed prob is 0.73

for $t=2$, filter prob is 0.50 $\}$ >
smoothed prob is 0.28

With smoothed analysis, student gets sleep derived more quickly as opposed to filtered analysis.

for $t=3$, filter prob is 0.10 $\}$ =
smoothed prob is 0.10

after $t=3$ they both converge to 0.10