CSCI 3202
Final Exam
Spring 2018

Name: ____

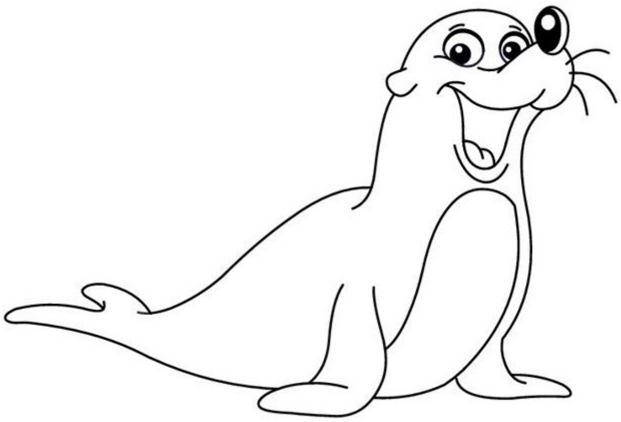
By writing my name here, I agree to the exam rules outlined below as well as in the course syllabus, and the CU Boulder Honor Code pledge:

On my honor, as a University of Colorado Boulder student, I have neither given nor received unauthorized assistance.

Read the following instructions.

- **Write your name** on the provided line above ****and on the back of the last page****.
- You may use a calculator provided that it cannot access the internet or store large amounts of data.
- You may **not** use a smartphone in any capacity.
- You must **clearly justify all conclusions** to receive full credit.
  A correct answer with no supporting work will receive little/no credit.
- If you need more space, there are some blank pages at the end of the exam.
  Please clearly indicate what work is associated with which problems.
- If you need to leave the exam room during the exam for any reason, **raise your hand** and show me that you **closed your exam** and **placed your phone** on top of it. Then, go do your thing.
- You have **50 minutes** for this exam.

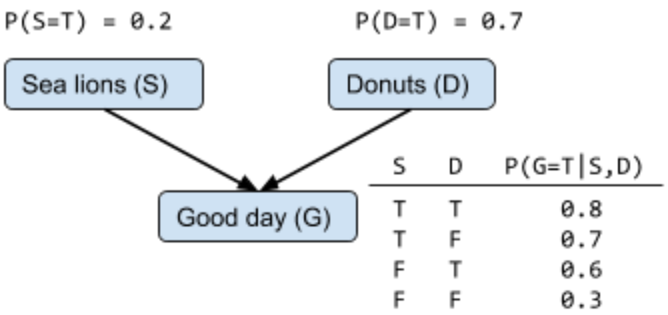
#	Points possible	Score
1	22	
2	12	
3	20	
4	20	
5	24	
For luck!	2	2
Total	100	



1. [22 points]

Consider the Bayesian network at right, which studies the relationship between exposure to sea lions, consumption of donuts, and having a good day. The variables are all Boolean (T/F) and are:

- S = did you see a sea lion that day
- D = did you eat a donut that day
- G = did you have a good day



S	D	P(G=T|S,D)
T	T	0.8
T	F	0.7
F	T	0.6
F	F	0.3

Please calculate the following probabilities. You do **not** need to fully simplify your answer; **obtaining an answer in terms of numbers and a normalizing constant is sufficient**.

a) $P(D = \mathrm{T} \mid G = \mathrm{T})$

b) $P(G = \mathrm{T}, S = \mathrm{F}, D = \mathrm{T})$

a) P(D = T | G = T) = C * P(G=T | D=T) P(D=T)

= C * Σ_s P(G=T | D=T, S=s) P(D=T) P(S=s)

= C * [P(G=T | D=T, S=T) P(D=T) P(S=T) + P(G=T | D=T, S=F) P(D=T) P(S=F)]

= C * [0.8 *0.2 * 0.7 + 0.6 * 0.8 *0.7]

= C * 0.448 ← this is sufficient

Doing the same for P(D = F | G = T) gives:

P(D = F | G = T) = C * [P(G=T | D=F, S=T) P(D=F) P(S=T) + P(G=T | D=F, S=F) P(D=F) P(S=F)]

= C * [0.7 *0.2 * 0.3 + 0.3 * 0.8 *0.3]

= C * 0.114

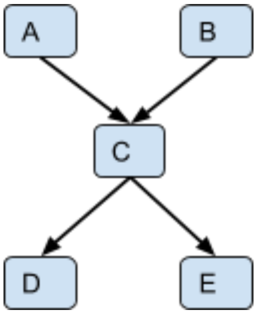
So: P(D=T | G=T) = 0.448 / (0.448 + 0.114) = 0.797

b) P(G=T, S=F, D=T) = P(G=T | S=F, D=T) * P(S=F) * P(D=T)

= 0.6 * 0.8 * 0.7

= 0.336

c) Now consider the Bayesian network given here, with variables A, B, C, D, and E. Which of the following relationships are implied by the structure of this Bayesian network? **Clearly** mark T or F (true or false) for each statement. No justification is needed.



F_____ $P(C \mid A) = P(C \mid A, B)$

T_____ $P(E \mid C, D) = P(E \mid C)$

F_____ $P(D \mid C) = P(D)$

T_____ $P(D \mid B, C) = P(D \mid C)$

2. [12 pts] Consider the data table below, which researchers painstakingly collected to study the relationship between exposure to sea lions, consumption of donuts, and having a good day. The variables are all Boolean (T/F), defined as in Problem 1a and b (other than that, these problems are unrelated).

The un-aligned formatting within the table is only so it is easier to tell which cells are T and which are F.

Provide reasonable estimates of the following, based on these data.

a) $P(S = \text{T})$

b) $P(G = \text{T} \mid S = \text{F}, D = \text{T})$

c) $P(G = \text{T}, S = \text{F}, D = \text{T})$

Sample #	**S**	**D**	**G**
1	T	T	T
2	T	T	F
3	T	F	T
4	T	F	T
5	F	T	T
6	F	T	T
7	F	T	F
8	F	T	F
9	F	F	F
10	F	F	F

a) Estimate of P(S=T) = # samples with S=T / # samples total
= 4/10 = 0.4

b) Estimate of P(G=T | S=F, D=T) = # samples with G=T, S=F, D=T / # samples with S=F, D=T
= 2/4 = 0.5

c) Estimate of P(G=T, S=F, D=T) = # samples with G=T, S=F, D=T # samples total
= 2/10 = 0.2

3. [20 pts]

a) Given this Markov model and the prior distribution [$P(X_0 = \mathrm{T}) = 0.5$, $P(X_0 = \mathrm{F}) = 0.5$], what is the probability distribution $P(X_1)$?

Potentially helpful suggestion: Remember that you can represent the prior distribution as:

$$P(X_0) = \begin{pmatrix} P(X_0 = T) \\ P(X_0 = F) \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Markov transition model:

X_t	$P(X_{t+1} \mid X_t)$
T	0.9
F	0.5

Use the "mini-forward" algorithm (Law of Total Probability on steroids!):

$$\begin{aligned} P(X_1) &= \sum_{x_0} P(X_1|x_0)P(x_0) \\ &= \begin{pmatrix} P(X_1 = T \mid X_0 = T)P(X_0 = T) + P(X_1 = T \mid X_0 = F)P(X_0 = F) \\ P(X_1 = F \mid X_0 = T)P(X_0 = T) + P(X_1 = F \mid X_0 = F)P(X_0 = F) \end{pmatrix} \\ &= \begin{pmatrix} 0.9 \cdot 0.5 + 0.5 \cdot 0.5 \\ 0.1 \cdot 0.5 + 0.5 \cdot 0.5 \end{pmatrix} \\ &= \begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix} \end{aligned}$$

b) Suppose this Markov model is a *hidden* one, and you can only make observations E_n such that the sensor model is given below. If we observe E_1=T, then what is our **unnormalized** filtered estimate $P(X_1 \mid E_1$=T)? ***Hint:*** *you could use your answer from part (a) here.*

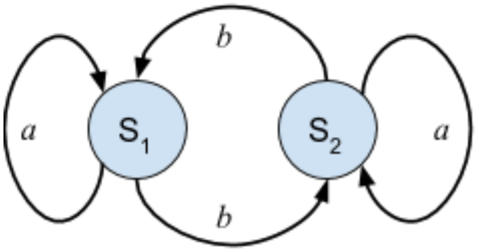
X_t	$P(E_t \mid X_t)$
T	0.8
F	0.4

Solution:

$$\begin{aligned} P(X_1 \mid E_1) &= \frac{P(E_1 \mid X_1)P(X_1)}{P(E_1)} = \alpha\ P(E_1 \mid X_1)P(X_1) \\ &= \alpha \begin{pmatrix} P(E_1 = T \mid X_1 = T)P(X_1 = T) \\ P(E_1 = T \mid X_1 = F)P(X_1 = F) \end{pmatrix} \\ &= \alpha \begin{pmatrix} 0.8 \cdot 0.7 \\ 0.4 \cdot 0.3 \end{pmatrix} \\ &= \alpha \begin{pmatrix} 0.56 \\ 0.12 \end{pmatrix} \end{aligned}$$

4. [20 pts]

Suppose a Markov decision process has two states, S_1 and S_2, in the state space. Action a does nothing, and action b takes the agent from whatever state it is in, to the other state. The actions are deterministic (results occur with probability 1). This is depicted in the figure to the right. Use:



- Discount factor $\gamma = 0.5$.
- Reward of state $S_1 = R(S_1) = 3$
- Reward of state $S_2 = R(S_2) = 2$

a) Start with the policy $\pi(S_1) = \pi(S_2) = a$. After one iteration of policy iteration, what are the estimates of the utilities of each state, $U(S_1)$ and $U(S_2)$, under the proposed policy?

Hint: Recall that the **policy evaluation** step under policy π amounts to solving a linear system for $U(S_1)$ and $U(S_2)$.

$U_0(S_1) = R(S_1) + \gamma\, U_0(S_1) = 3 + 0.5 * U_0(S_1)$
$\Rightarrow \frac{1}{2}\, U_0(S_1) = 3$
$\Rightarrow \boldsymbol{U_0(S_1) = 6}$

And

$U_0(S_2) = R(S_2) + \gamma\, U_0(S_2) = 2 + 0.5 * U_0(S_2)$
$\Rightarrow \frac{1}{2}\, U_0(S_2) = 2$
$\Rightarrow \boldsymbol{U_0(S_2) = 4}$

b) What happens in the **policy improvement** step of this first iteration of policy iteration? Select all that apply. No justification is needed.

_____ nothing   _____ $\pi(S_1)$ is set to b __X___ $\pi(S_2)$ is set to b

5. [24 pts (4 each)] **Multiple choice.** Circle the letter corresponding to your answer for each problem.

[1] Suppose you are using Adaptive Dynamic Programming (ADP) to learn the transition model *P*(*s*' | *s*, *a*) for a Markov decision process. You count in the dictionaries `N`, `Nsa` and `Ntsa` the following:

- `N[s]` = number of times state *s* has been visited
- `Nsa[(s,a)]` = number of times action *a* has been taken from state *s*
- `Ntsa[(s',s,a)]` = number of times action *a* has been taken from state *s* and led to subsequent state *s'*

Which of the following is the most appropriate estimate of *P*(*s*' | *s*, *a*)?

a. `Ntsa[(s',s,a)] / N[s]`

b. **`Ntsa[(s',s,a)] / Nsa[(s, a)]`**

c. `Nsa[(s,a)] / Ntsa[(s', s, a)]`

d. `Nsa[(s,a)] / N[s]`

[2] Which of the following is the best example of the concept of Greedy in the Limit of Infinite Exploration (GLIE)? Recall that an ε-greedy agent picks a random move with probability ε, and a greedy move with probability 1-ε.

a. An ε-greedy agent where ε = 0.1 is fixed.

b. An ε-greedy agent where ε maximizes the current best estimate of expected utility.

c. An ε-greedy agent where ε decreases as the number of training episodes increases.

d. An ε-greedy agent where ε increases as the number of training episodes increases.

[3] Suppose you use value iteration to solve for the optimal utilities for every state in your Markov decision process model. Now, how do you determine the optimal policy for each state?

a. The optimal policy is the action that maximizes the expected reward of the subsequent state.

b. The optimal policy is the action that maximizes the utility of the most likely subsequent state.

c. The optimal policy is the action that maximizes the minimum utility of the subsequent state.

d. The optimal policy is the action that maximizes the expected utility of the subsequent state.

[4] In the Bayesian sock count estimation problem from class, we used an accept/reject algorithm that would assign a probability of 1 to any simulation that matches our data, and 0 if the simulation did not match our data. What component of the Bayesian statistical framework was that algorithm/function a part of?

a. prior distribution

b. likelihood function

c. posterior distribution

d. evidence

[5] In Problem 1a, how would your answer for $P(D = \mathrm{T} \mid G = \mathrm{T})$ change if you also knew that $S = \mathrm{T}$?

a. It would decrease

b. It would increase

c. It would stay the same

+2 extra credit: What is this phenomenon called? _____**Explaining away**____________________

[6] Suppose the following diagram depicts the optimal utilities for a Markov decision process. The available actions are to move North, South, East or West. Suppose the transition model is:

- If the agent tries to move in a particular direction, there is 1/3 probability they make the move successfully, 1/3 probability they end up in the tile to their right, and 1/3 probability they end up in the tile to their left.
- For example, if the agent executed action North from (2, 1), the possible results and their probabilities are:
  - with probability 1/3, the agent would be in (2, 2)
  - with probability 1/3, the agent would be in (1, 1)
  - with probability 1/3, the agent would be in (3, 1)
- You do not need to know any other information to answer this question.

What is the optimal policy for the state (2, 2)? ((2, 2) is the lightly shaded state.)

a. North ← Also accepted (ambiguous)

b. South

c. East

d. West

3	**1**	**12**	**8**
2	**9**	**5**	**3**
1	**2**	**6**	**4**
	1	2	3