CMPSCI 240: Reasoning Under Uncertainty
Third Midterm Exam

April 19, 2017.

Name: _____ ID: _____

Instructions:

- Answer the questions directly on the exam pages.
- Show all your work for each question. Providing more detail including comments and explanations can help with assignment of partial credit.
- *Unless the question specifies otherwise,* you may give your answer using arithmetic operations, such as addition, multiplication, "choose" notation and factorials (e.g., "$9 \times 35! + 2$" or "$0.5 \times 0.3/(0.2 \times 0.5 + 0.9 \times 0.1)$" is fine).
- If you need extra space, use the back of a page.
- No books, notes, calculators or other electronic devices are allowed. Any cheating will result in a grade of 0.
- If you have questions during the exam, raise your hand.
- The formulas for some standard random variables can be found on the last page.

Question	Value	Points Earned
1	14	
2	10	
3	10+2 (Extra Credit)	
4	10	
5	10	
Total	54+2 (Extra Credit)	

Question 1. (*14 points*) Indicate whether each of the following statements is TRUE or FALSE. No justification is required.

1.1 (*2 points*): *If I throw four balls and each is equally likely to land in any one of three bins then the probability that they all land in different bins is* 0.

1.2 (*2 points*): *If X is a binomial random variable then $E(X) \geq var(X)$.*

1.3 (*2 points*): *The expected number of times you need to roll a normal six-sided dice until you've seen all sides is* $6/6 + 6/5 + 6/4 + 6/3 + 6/2 + 6/1$.

1.4 (*2 points*): *If X and Y are independent random variables and each takes the values 0 or 1, then the joint distribution of X and Y is fully determined by the values of $P(X = 0)$ and $P(Y = 1)$.*

1.5 (*2 points*): *Suppose you have two hypotheses H_1 and H_2 and you observe event D has occurred. Then $P(D|H_1) \geq P(D|H_2)$ implies that H_1 is the maximum a posteriori (MAP) hypothesis.*

1.6 (*2 points*): *$E(X + Y) = E(X) + E(Y)$ implies X and Y are independent.*

1.7 (*2 points*): *If 10 students are each equally likely to be born in any of the 12 months in the year, then the expected number of students born in January is* $10/12$.

Question 2. (*10 points*) Three students are taking a class that has five discussion sections at different times in the day. Each student is equally likely to pick any of the sections and they each pick independently. *To get full marks on this question, you should simplify your answers fully.*

2.1 (*2 points*): *What's the probability they are all in the same discussion section?*

2.2 (*2 points*): *What's the probability they are all in different discussion sections?*

2.3 (*2 points*): *What's the probability exactly two students are in the earliest discussion section?*

2.4 (*2 points*): *What's the probability there exists a section with exactly two of these students?*

2.5 (*2 points*): *What's the probability the third student picks a discussion section that doesn't include either of the first two students?*

Question 3. (*12 points*) A shop sells two different types of six-sided dice, a *normal* dice and a *weird* dice. The normal dice is equally likely to show the numbers 1, 2, 3, 4, 5, and 6 whereas the weird dice is equally likely to show the numbers 1, 2, 3, 4, and 5 but never shows the value 6.

You're not sure which type of dice you bought: with probability 2/3 it was a normal dice (call this event N) and with probability 1/3 it was a weird dice (call this event W). You roll it twice and let X be the number shown on the first roll and let Y be the number shown on the second roll.

3.1 (*4 points*): *What are the values of the following probabilities:*

$P(X = 1|N) =$ $P(X \leq 2|N) =$ $P(X = 6|W) =$ $P(X \leq 2|W) =$

3.2 (*2 points*): *If $X \leq 5$, is it more likely you have the normal dice or the weird dice?*

3.3 (*2 points*): *If $X \leq 5$ and $Y \leq 5$, is it more likely you have the normal dice or the weird dice?*

3.4 (*2 points*): *Are X and Y independent? Justify your answer.*

3.5 (*2 points*): **Extra Credit:** *What's the expected number of rolls of the normal dice until you've seen all odd numbers of the dice?*

Question 4. (*10 points*) There has been some research on DNA sequences that model them as Markov chains. For the purposes of this question all you need to know about DNA is that it is a random sequence involving the characters $\texttt{A}, \texttt{G}, \texttt{C}, \texttt{T}$, e.g., $\texttt{AGCTAT}\ldots$. For $i \in \{1, 2, 3, \ldots\}$, let X_i be ith letter of the random sequence, e.g., in the previous example $X_3 = \texttt{C}$. Suppose the researchers modeled the random DNA sequence as a Markov chain with states $\texttt{A}, \texttt{G}, \texttt{C}, \texttt{T}$ where, for each $i \geq 2$:

$$P(X_i = \texttt{C} | X_{i-1} = \texttt{G}) = 1 \ , \quad P(X_i = \texttt{T} | X_{i-1} = \texttt{C}) = 1 \ , \quad P(X_i = \texttt{A} | X_{i-1} = \texttt{T}) = 1 \ ,$$

$$P(X_i = \texttt{G} | X_{i-1} = \texttt{A}) = 1/2 \ , \quad P(X_i = \texttt{T} | X_{i-1} = \texttt{A}) = 1/2$$

and all other transition probabilities are zero. For example, if the first character is known to be $\texttt{T}$ then the probability that the first five characters are $\texttt{TATAG}$ is $1 \times 1/2 \times 1 \times 1/2 = 1/4$.

4.1 (*2 points*): *Draw the transition diagram for this Markov chain.* **Hint:** *There should be four nodes and remember to include the transition probabilities.*

4.2 (*4 points*): *Suppose the first character is known to be* $\texttt{A}$.

- *What is the probability the second character is* $\texttt{T}$*?*

- *What is the probability the first five characters are* $\texttt{ATATA}$*?*

- *What is the probability that the fifth character is* $\texttt{A}$*?*

- *What is the expected position of the first* $\texttt{T}$ *in the sequence?*

4.3 (*2 points*): *Suppose you're a biologist and your equipment for reading DNA sequences fails to identify the first character of a sequence and without any more information, you assume it is equally likely to be any of the four characters. However, the equipment does tell you that* $X_2 = \texttt{T}$. *Given this information, which character is most likely to be at the start of the sequence.*

4.4 (*2 points*): *What is the steady state of the distribution of the Markov chain?*

Question 5. (*10 points*) *To get full marks on this question, you should simplify your answers fully.* Suppose that a patient is experiencing back problems. This could be because of the new chair he bought that, although it looked cool, might actually be badly designed or it could be because of a sports injury.

- Let $C = 1$ if his office chair is badly designed and $C = 0$ otherwise.
- Let $T = 1$ if he was playing tennis last weekend and $T = 0$ otherwise.
- Let $B = 1$ if he has damaged his back and $B = 0$ otherwise.

You may assume that C and T are independent random variables and that

$$P(C = 1) = 1/2\ , \qquad P(T = 1) = 1/2\ , \qquad P(B = 1|C = 0, T = 0) = 0\ ,$$

$$P(B = 1|C = 0, T = 1) = 1/4\ , \qquad P(B = 1|C = 1, T = 0) = 1/2\ , \qquad P(B = 1|C = 1, T = 1) = 1$$

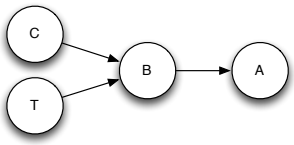
5.1 (*4 points*): *What are the values of the following:*

$$\begin{aligned} P(C = 1, T = 1) &= \\ P(C = 1, T = 0) &= \\ P(C = 1|C = 1, T = 1) &= \\ var(CT) &= \end{aligned}$$

5.2 (*4 points*): *What are the values of the following probabilities:*

$$\begin{aligned} P(B = 1, C = 1, T = 1) &= \\ P(B = 1) &= \\ P(B = 1, C = 1) &= \\ P(B = 1|C = 1) &= \end{aligned}$$

Let $A = 1$ if the patient's back aches and $A = 0$ otherwise. Suppose the joint distribution of C, T, B, and A can be modeled by the following Bayesian network.



5.3 (*2 points*): *In addition to the 6 conditional probabilities given at the start of the question, what two other conditional probabilities are sufficient to specify the joint distribution of C, T, B, A?*

Standard Random Variables

- **Bernoulli Random Variable with parameter** $p \in [0,1]$:

$$P(X=k) = \begin{cases} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{cases} \quad , \quad E(X) = p \quad , \quad var(X) = p(1-p)$$

- **Binomial Random Variable with parameters** $p \in [0,1]$ **and** $N \in \{1,2,3,\ldots\}$:

$$\text{For } k \in \{0,1,2,\ldots,N\} : P(X=k) = \binom{N}{k} p^k (1-p)^{N-k} \ , \ E(X) = Np \ , \ var(X) = Np(1-p)$$

- **Geometric Random Variable with parameter** $p \in [0,1]$:

$$\text{For } k \in \{1,2,3,\ldots\} : \ P(X=k) = (1-p)^{k-1} \cdot p \quad , \quad E(X) = \frac{1}{p} \quad , \quad var(X) = (1-p)/p^2$$

- **Poisson Random Variable with parameter** $\lambda > 0$:

$$\text{For } k \in \{0,1,2,\ldots\} : \ P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad , \quad E(X) = \lambda \quad , \quad var(X) = \lambda$$

- **Discrete Uniform Random Variable with parameters** $a, b \in \mathbb{Z}$ **and** $a < b$:

$$\text{For } k \in \{a, a+1, \ldots b\} : \ P(X=k) = \frac{1}{b-a+1} \ , \ E(X) = \frac{a+b}{2} \ , \ var(X) = \frac{(b-a+1)^2 - 1}{12}$$

Bayes Formula

- If $A_1, \ldots, A_n$ partition Ω then for any event B:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$