CMPSCI 240: Reasoning Under Uncertainty
Third Midterm Exam

April 14, 2016.

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	10	
2	10	
3	10	
4	10	
5	8	
6	10+2 (Extra Credit)	
Total	58+2 (Extra Credit)	

Question 1. (*10 points*) Indicate whether each of the following statements is TRUE or FALSE. No justification is required.

1.1 (*2 points*): *For any random variables X, Y, and Z, and values a, b, and c,*

$$P(X = a, Y = b, Z = c) = P(X = a)P(Z = c \mid X = a)P(Y = b \mid X = a, Z = c)$$

1.2 (*2 points*): *For any nonnegative random variable X, $P(X \geq 2E[X]) \leq \frac{1}{2}$.*

1.3 (*2 points*): *For any two random variables X and Y, $Var(X+Y) = Var(X) + Var(Y)$.*

1.4 (*2 points*): *Consider a Markov Chain with two states S and T and the state transition matrix*

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The above Markov Chain has a unique stationary distribution.

1.5 (*2 points*): *Suppose you throw 2 balls and each ball is equally likely to land in any one of 10 bins, then the probability that the two balls do* **not** *lie in the same bin is 0.1.*

Question 2. (*10 points*) Consider the Markov chain with four states and transition matrix:

$$A = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

2.1 (*2 points*): *Draw the transition diagram.*

2.2 (*2 points*): *Circle the two words that apply to the above Markov Chain:*

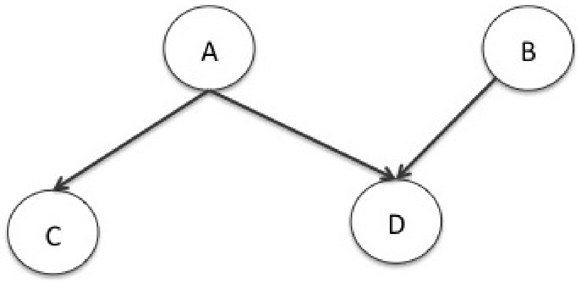
aperiodic *periodic* *reducible* *irreducible*

2.3 (*2 points*): *Suppose that the Markov chain is initially in state one. What is the probability that it is in state four after four steps? Show your working.*

2.4 (*2 points*): *Suppose that the Markov chain is initially in state one. What is the probability that it is in state four after four steps, and state two after six steps? Show your working.*

2.5 (*2 points*): *Find the steady state distribution. Show your working.*

Question 3. (*10 points*) In the following Bayesian network, A, B, C, D are binary random variables.



3.1 (*2 points*): *What are the factors of the above Bayesian network?*

3.2 (*2 points*): *What is the storage space required to represent the joint distributions on* A, B, C, D*? Show your working.*

3.3 (*3 points*): *Suppose you are given the following probabilities* $P(A = 0) = P(B = 1) = 0.9$, $P(C = 1|A = 0) = 0.2$, $P(C = 1|A = 1) = 0.4$. *Compute the following probabilities.*

$$P(C = 1) =$$

$$P(C = 1, B = 1) =$$

$$P(C = 1, B = 1 \mid A = 1) =$$

3.4 (*3 points*): *We are further given that* $P(D = 1 \mid A = x, B = y) = 0.3$ *where* $x = 0$ *or* 1, *and* $y = 0$ *or* 1. *Compute the following probabilities.*

$$P(D = 1) =$$

$$P(D = 0) =$$

$$P(A = 0 \mid D = 1) =$$

Question 4. (*10 points*) Diabetes mellitus is a chronic disease and a major public health challenge worldwide. Among the 1000 patients that were tested for diabetes in the last month, 400 were found to have diabetes. Among the patients that were found to have diabetes, 200 of them have age above 40. Among the patients that were tested not to have diabetes, 300 of them have age above 40. Among the patients that were tested to have diabetes, 300 of them have prior incidence of diabetes in the family. Among the patients that were tested not to have diabetes, 100 of them have prior incidence of diabetes in the family.

Let A and B be the events that a new patient has age above 40 and has prior incidence of diabetes in the family respectively. We have two hypotheses:

H_1 = new patient has diabetes and H_2 = new patient does not have diabetes

You may assume that $P(A \cap B \mid H_1) = P(A \mid H_1)P(B \mid H_1)$, and similarly $P(A \cap B \mid H_2) = P(A \mid H_2)P(B \mid H_2)$.

4.1 (*2 points*): *What values should you use for the priors:*

$P(H_1) =$ $P(H_2) =$

4.2 (*2 points*): *What values should you use for the likelihoods:*

$P(A \mid H_1) =$ $P(A \mid H_2) =$

$P(B \mid H_1) =$ $P(B \mid H_2) =$

4.3 (*2 points*): *What is the maximum a posteri (MAP) hypotheses if the new patient has age* 45 *and his grandparent had diabetes? Show your working.*

4.4 (*2 points*): *Given that A and B are conditionally independent on H_1, is it necessarily true that A^c and B^c are conditionally independent on H_1? Justify your answer.*

4.5 (*2 points*): *What is the maximum a posteri (MAP) hypotheses if the new patient has age* 30 *and has no history of diabetes in the family? Show your working.*

Question 5. (*8 points*) Imagine you're running an online store and lately you've been overrun with fraudulent orders. You estimate that about 10% of all orders coming in are fraudulent. You have noticed that fraudulent orders often use gift card and multiple promotional codes.

Let G and M be the events that an order uses gift card, and on order uses multiple promotional codes respectively. Let F be the event that an order is fraudulent, and N be the event that an oder is not fraudulent.

5.1 (*3 points*): *Lets say that gift card is used in* 20% *of the total number of (fraudulent and non fraudulent) orders, and based on your research, among the fraudulent orders* 60% *use gift cards. What values should you use for the following probabilities?*

$P(F = 1) =$ $\qquad\qquad$ $P(N = 1) =$

$P(G = 1) =$ $\qquad\qquad$ $P(G = 0) =$

$P(G = 1 \mid F = 1) =$ $\qquad\qquad$ $P(G = 0 \mid F = 1) =$

$P(F = 1 \mid G = 1) =$

5.2 (*3 points*): *Suppose among the fraudulent orders* 50% *of them use multiple promo codes, whereas, among the non-fraudulent orders only* 30% *uses multiple promo codes. Compute the followings.*

$P(M = 1 \mid F = 1) =$ $\qquad\qquad$ $P(M = 1 \mid N = 1) =$

$P(G = 1, N = 1) =$

$P(G = 1 \mid N = 1) =$

5.3 (*2 points*): *You get a new order which uses a gift card and uses multiple promo codes. What is the maximum a posteri (MAP) hypothesis using Naive Bayes classifier? Show your working.*

Question 6. (*12 points*) Suppose you throw 100 balls into 100 bins.

6.1 (*2 points*): *Let B_i be the number of balls that land in the ith bin. What is the value of $P(B_i = 5)$?*

6.2 (*2 points*): *What is the value of $E(B_i)$ and $var(B_i)$?*

6.3 (*2 points*): *Let $B = \max(B_1, B_2, \ldots, B_n)$ be the number of balls in the bin with the most balls. What is the value of $P(B \geq 1)$?*

6.4 (*2 points*): *What is the value of $P(B \geq 2)$?*

6.5 (*2 points*): *Recall from the class $P(B_i \geq k) \leq (m/n)^k/k!$. What is the value of $P(B \geq 15)$?*

6.6 (*2 points*): **Extra credit.** *What is the value of $E(B_i + B_j)$ and $var(B_i + B_j)$ where $i \neq j$?*

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** $N \in \{1,2,3,\ldots\}$:

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** $p \in [0,1]$:

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

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

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

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

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