Due: September 22nd , 2015 (Fri.), 5 PM   **Instructor: Dr. M. E. Kim**

Assignment 1 (150 points.)

Instruction:

For each question, you can *manually compute* it or write a *program* to compute it.

If you manually compute it,

1) Define the variables and their values: e.g.) C = Cancer - a Boolean Random Variable (R.V), T = Test –a domain of T = { +, – } for a positive result or for a negative result, respectively, etc.
2) Define the given probabilities with the variables and the values: e.g.) P(c) = 0.5
3) Define what probability is asked to compute in the question: e.g.) P(-|c) – a probability that a patient had a negative test result though a patient has a cancer.
4) Show the essential steps of computation with the formulas to get the final solution.
   e.g*.) P(A | B) = P(A, B)/P(B) = P(B|A)P(A)/ P(B)* = 0.3 * 0.5 / 0.8 = 1/6.

   Do NOT simply describe your computations verbally without the proper definitions of probabilities.

If you program it,

1) Run your program to get the answer.
2) Capture the screen image of your answer.
3) Insert the image to the corresponding question in the HW file.

Any solution neither with the proper formulas nor with the computational steps will get **no** point.

You should work on the assignment**, *independently*.** – Any plagiarism or collaboration is **not** allowed. – Refer to the syllabus and Code of Student Life.
http://und.edu/student-affairs/code-of-student-life/_files/codepdfs/appendix/iiia/iiia-3.pdf

Submission:

1. Prepare your homework using a word processor, a MS Word: no **.pdf** file is acceptable.
2. Write your name in the homework.
3. Name your file in the format**, HW1-YourLastName.docx**: e.g.) HW1-Kim.docx
4. If you have files of source codes, (a) create a directory being named as 'HW1-YourLastName', (b) locate every file including the HW file in the directory and (c) compress the directory to .zip file.
5. Upload your HW file or the compressed .zip directory in the blackboard system.

Q1. [20] **Cancer Diagnosis with One Test.**

After the lab test, the doctor has bad news and good news to a patient.
The bad news is that he tested positive(+) for a cancer and that the accuracy of test is 90% (i.e. the probability of testing positive(+) when he does have a cancer is 0.90) while the specificity of test is 80% as is the probability of testing negative(-) when he doesn't have a cancer. The good news is that this is a rare form of cancer, striking only 1 % of population has this type of cancer.

Using the following variables, compute the following probability.

        C : the patient has a particular form of cancer or not $\in$ {c, ¬c }

        T_1 : a lab test T_1 with two possible outcomes $\in$ {+, - }

              + : positive result,   - : negative result.

1) [10] Compute the probability that a patient has a cancer when he has a positive test result.
2) [10] Compute the probability that a patient doesn't have a cancer when he has a negative test result.

Q2. [30] **Cancer Diagnosis with Two Tests.**

In Q1, the 2nd lab test (T_2) is given to the patient as well as the 1st test (T_1) for a more accurate result. For the 2nd test, both its accuracy and specificity are as same as the 1st test.

1) [10] Compute the probability that he has a cancer when both test results are positive.
    i.e. T_1= +, T_2= +.
2) [10] Compute the probability that he has a cancer when T_1 = + but T_2 = - .
3) [10] Compute the probability that test 2 reveals positive when test 1 gave positive result.

Q3. [10] Bayesian Network Construction

Suppose we want to use the diagnostic assistant to diagnose whether there is a fire in a building based on noisy sensor information and possibly conflicting explanations of what could be going on. The agent receives a report about whether everyone is leaving the building. Suppose the report sensor is noisy: It sometimes reports leaving when there is no exodus (a false positive), and it sometimes does not report when everyone is leaving (a false negative). Suppose the fire alarm going off can cause the leaving, but this is not a deterministic relationship. Either tempering or fire could affect the alarm. Fire also causes smoke to rise from the building.

Construct the Bayesian network using the following Boolean variables in the following order:
- *Tampering* is true when there is tampering with the alarm.
- *Fire* is true when there is a fire.
- *Alarm* is true when the alarm sounds.
- *Smoke* is true when there is smoke.
- *Leaving* is true if there are many people leaving the building at once.
- *Report* is true if there is a report given by someone of people leaving. *Report* is false if there is no report of leaving.

Q4. [30 pts.] Inference in Bayesian Network

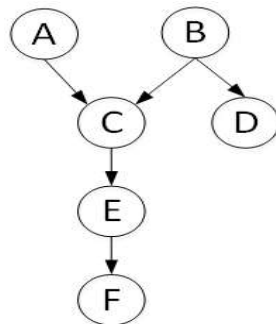Consider the simple Bayes net below with Boolean variables:
$B = BrokeElectionLaw,\ I = Indicted,\ M = PoliticallyMotivatedProsecutor,\ G = FoundGuilty,\ J = Jailed.$

B	M	P(I)
t	t	.9
t	f	.5
f	t	.5
f	f	.1

P(B)
.9

P(M)
.1

B → I ← M
I → G
G → J

B	I	M	P(G)
t	t	t	.9
t	t	f	.8
t	f	t	.0
t	f	f	.0
f	t	t	.2
f	t	f	.1
f	f	t	.0
f	f	f	.0

G	P(J)
t	.9
f	.0
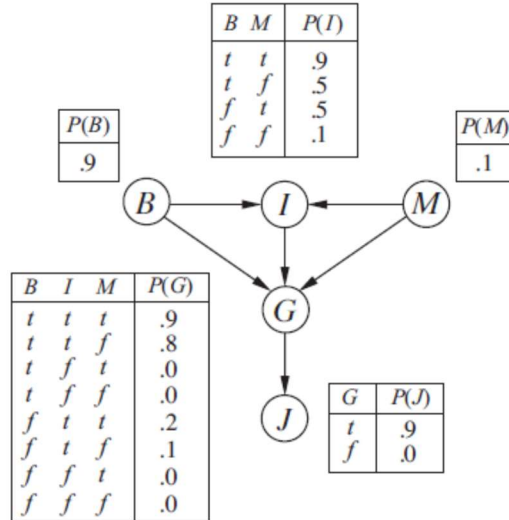
1. [10] Which of the following are asserted by the network structure? Explain.
   (a) P(B, I, M) = P(B)P(I)P(M).
   (b) P(J|G) = P(J|G,I)
   (c) P(M|G,B,I) = P(M|G,B,I,J)
2. [10] Calculate the value of P(b ,I ¬m, g, j). Show the essential steps of your computation.
3. [10] Calculate the probability that someone goes to jail given that they broke the law, have been indicated, and face a politically motivated prosecutor. Define the probability that you have to compute and show the essential computational steps.

Q5. [30 pts.] Inference with Variable Elimination

Consider the following belief network with Boolean variables and the following conditional probabilities:

A → C
B → C
B → D
C → E
E → F

P(a) = 0.02   P(d | b) = 0.9
P(b) = 0.01   P(d | ¬b) = 0.01
P(c | a, b) = 0.5   P(e | c) = 0.88

$P(c \mid a, \neg b) = 0.99$ $P(e \mid \neg c) = 0.001$
$P(c \mid \neg a, b) = 0.85$ $P(f \mid e) = 0.75$
$P(c \mid \neg a, \neg b) = 0.0001$ $P(f \mid \neg e) = 0.01$

(1) [15] Compute $P(A \mid d, f)$ using Variable Elimination (VE). Which variables are irrelevant to inference? First, prune irrelevant variables. Show the complete computation steps with the factors that are created for a given elimination ordering.

(2) [15] c) Compute $P(e \mid b)$ using VE. a) Which variables are irrelevant to inference? b) Which factors can be reused from (1)? d) Show the factors that are different from those in (1).

Q6. [30 pts.] Inference in Bayesian Network

In the given Bayesian network, compute the following probabilities.

(1) [15] What is the probability that a middle-aged male patient who is over 40 but up to 60 gets malignant cancer?
(2) [15] What is the probability that a middle-aged male patient with a lung tumor who is over 40 but up to 60 to have malignant cancer?

Age	P(age)
0 ≤ age < 20	.2
20 ≤ age < 40	.25
40 < age ≤ 60	.25
60 < age	.3

Gender	P(gender)
Male	.6
Female	.4

Age	P(exposure to Toxics | Age)
0 ≤ age < 20	.1
20 ≤ age < 40	.3
40 < age ≤ 60	.6
60 < age	.7

Age	Gender	P(smoking|Age, Gender)
0 ≤ age < 20	Male	.05
0 ≤ age < 20	Female	.02
20 ≤ age < 40	Male	.4
20 ≤ age < 40	Female	.3
40 < age ≤ 60	Male	.4
40 < age ≤ 60	Female	.2
60 < age	Male	.3
60 < age	Female	.1

Exposure to Toxics	Smoking	P(Cancer=malignant | Exposure, Smoking)
True	True	.9
True	False	.6
False	True	.6
False	false	.1

Cancer	P(lung tumor|Cancer)
malignant	.99
benign	.4

Cancer	P(Serum Calcium=high|Cancer)	P(Serum Calcium=med|Cancer)
malignant	.6	.3
benign	.1	.4