Master's Degree on Foundations of Data Science

Probabilistic Graphical Models (PGM)

June 16th, 2021

Universitat de Barcelona

Full name: _____

- Include the reasoning steps followed to reach to the result.
- Each question scores up to 1 point. The final score is:

	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6	Q#7	Q#8	Q#9	Q#10	**Total**
Score											

Question #1: Basic probability.

Calculate the probability of **really** having Covid given a Positive test, $p(C = +|T = +)$, if we know that the prevalence of the illness (marginal probability of having COVID) is 0.05 and the sensitivity of the test is 80% (with, unfortunately, a 30% of false positives),

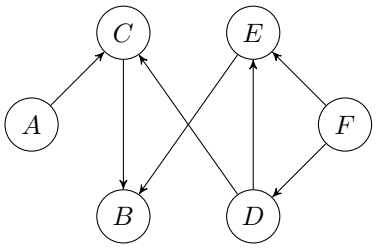
C	$p(C)$
-	
+	

T	C	$p(T\|C)$
-	-	
+	-	
-	+	
+	+	

(____/2)

Question #2: Independencies in Bayesian networks.

Which of the following conditional independence statements are true? (**Multiple choice**)



[ ] $A \perp\!\!\!\perp D|B$

[ ] $C \perp\!\!\!\perp E|D$

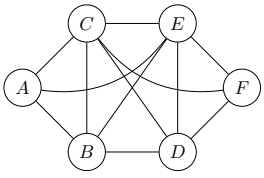
[ ] $A \perp\!\!\!\perp D|C$

[ ] $C \perp\!\!\!\perp F|D$

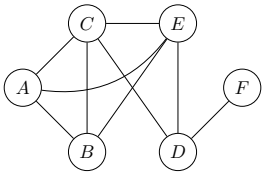
(____/4)

Question #3: Variable Elimination.

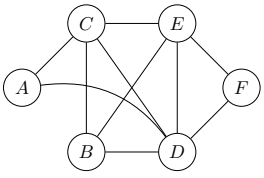
Considering the previous structure G, which of the following structures is the induced graph resulting from following the ordering (A, F, C, D, B, E)? (**Single choice**)



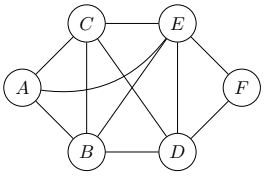
[ ] G_1



[ ] G_2



[ ] G_3



[ ] G_4

(____/1)

Question #4: Factorization in Bayesian networks (I-map).

Consider the previous structure G and the set of probability distributions, P, which factorize as displayed. Which distribution P is guaranteed to factorize according to G $[I(G) \subseteq I(P)]$? (**Multiple choice**)

[ ] $P(A, B, C, D, E, F) = P(A)P(B)P(C)P(D)P(E)P(F)$

[ ] $P(A, B, C, D, E, F) = P(A)P(B|C, E)P(C|A, D)P(D|F)P(E|D, F)P(F)$

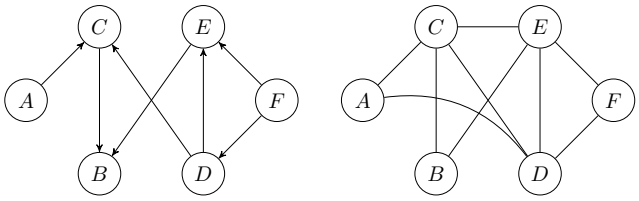
[ ] $P(A, B, C, D, E, F) = P(A, C)P(B|C, E)P(C|D)P(D|F)P(E|D, F)P(F)$

[ ] $P(A, B, C, D, E, F) = P(A)P(C|D, A)P(B|C, E)P(D, E, F)$

(____/4)

Question #5: Independencies in Markov networks.

Identify the independence statement(s) satisfied by the DAG of the BN on the left which is/are not satisfied by the undirected graph of the MN on the right. (**Multiple choice**)

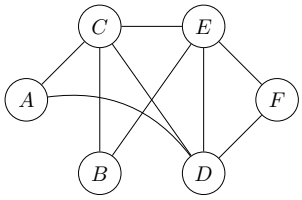


[ ] $A \not\perp\!\!\!\perp D|B$
[ ] $A \perp\!\!\!\perp F$
[ ] $C \perp\!\!\!\perp E$
[ ] $B \perp\!\!\!\perp D|C, E$

(____/4)

Question #6: Variable elimination.

Consider again the undirected graph, H, and the elimination order A, E, D, C, F, B. Which is the size (in terms of no. variables) of the largest intermediate factor found.



(____/2)

Question #7: Factorization in Markov networks.

Consider again the previous graph, H, and the set of probability distributions, P, which factorize as displayed. Which distribution P is guaranteed to factorize according to H? (**Multiple choice**)

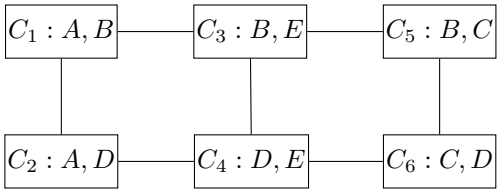
[ ] $P(A, B, C, D, E, F) \propto \phi_1(A, D, C) \times \phi_2(C, D, E) \times \phi_3(E, F)$
[ ] $P(A, B, C, D, E, F) \propto \phi_1(A, B, D, E) \times \phi_2(C, D) \times \phi_3(D, F)$
[ ] $P(A, B, C, D, E, F) \propto \phi_1(A, D, C) \times \phi_2(B, C, E) \times \phi_3(C, D, E) \times \phi_4(D, E, F)$
[ ] $P(A, B, C, D, E, F) \propto \phi_1(D, E, F) \times \phi_2(A, B, C)$

(____/4)

Question #8: Message Passing in a Cluster Graph.

Suppose we wish to perform inference in this Cluster Graph. Which expression correctly represents the message $\delta_{3\to4}$ that cluster C_3 sends to cluster C_4 in belief propagation? (**Single choice**)

Assume that the sepsets are composed of all the variables in the intersection of the adjacent cliques.



[ ] $\delta_{3\to4}(E) = \sum_B \psi_3(B, E) \cdot \delta_{1\to3}(B) \cdot \delta_{5\to3}(B) \cdot \delta_{4\to3}(E)$
[ ] $\delta_{3\to4}(E) = \sum_B \psi_3(B, E) \cdot \delta_{1\to3}(B) \cdot \delta_{5\to3}(B)$
[ ] $\delta_{3\to4}(B, E) = \psi_4(B, E) \cdot \delta_{1\to3}(B) \cdot \delta_{5\to3}(B)$
[ ] $\delta_{3\to4}(D, E) = \sum_{A,B,C} \delta_{1\to3}(A, B) \cdot \delta_{5\to3}(C, B)$

(____/1)

Question #9: Parametric learning.

Estimate the distribution of $P(Flu|Fever = yes, Cough = dry)$ from the following dataset with 30 instances.

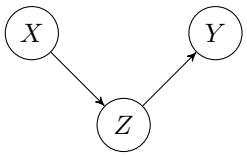
If we are trying to save space, which is the minimum number of parameters to encode CPD $P(Flu|Fever, Cough)$?

Cough	*Fever*	*Headache*	*Cold*	*Flu*
no	yes	no	no	yes
dry	yes	yes	no	no
prod	no	yes	yes	no
dry	yes	yes	no	no
no	yes	no	yes	no
no	no	no	yes	no
prod	no	yes	yes	yes
no	yes	yes	no	yes
no	yes	no	no	yes
prod	yes	no	yes	yes
prod	no	no	no	yes
dry	no	yes	no	yes
dry	yes	no	no	yes
no	yes	yes	yes	yes
no	no	no	yes	no
prod	yes	yes	yes	no
no	no	yes	no	no
dry	yes	no	yes	yes
dry	yes	yes	yes	no
no	yes	yes	yes	yes
prod	no	yes	yes	yes
no	no	yes	no	yes
prod	no	yes	yes	no
dry	yes	no	yes	yes
no	yes	no	no	yes
dry	yes	yes	yes	no
prod	no	yes	yes	no
prod	yes	yes	no	no
no	yes	no	no	no
no	yes	yes	no	yes

(____/4)

Question #10: Computing Sufficient Statistics.

Given this BN and this dataset, which is the value of the sufficient statistic $\tilde{M}[y_1, z_0]$?



X	$P(X)$
x_0	0.6

Z	X	$P(Z\|X)$
z_0	x_0	0.7
z_0	x_1	0.15

Y	Z	$P(Y\|Z)$
y_0	z_0	0.55
y_0	z_1	0.7

dataset
$(x_1, ?, z_0)$
$(x_0, y_1, ?)$
(x_1, y_0, z_0)
$(?, y_1, z_0)$
$(x_1, y_0, ?)$
(x_0, y_1, z_0)
$(x_0, ?, z_0)$
$(x_1, y_1, ?)$
(x_1, y_0, z_1)
$(x_1, y_0, ?)$
$(?, y_0, z_0)$
(x_0, y_0, z_0)

(____/4)