Master's Degree on Foundations of Data Science
Probabilistic Graphical Models (PGM)
May 15th, 2019

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 the weighted sum according to:

	Q#1	Q#2	Q#3	Q#4	Q#5	Q#6	Q#7	Q#8	Q#9	Q#10	Q#11	Q#12	Q#13	**Total**
Value	$0.0\dot{6}$	$0.0\dot{6}$	$0.0\dot{6}$	$0.0\dot{6}$	0.10	0.10	$0.0\dot{6}$	0.10	$0.0\dot{6}$	$0.0\dot{6}$	$0.0\dot{6}$	$0.0\dot{6}$	0.10	(to 10)
Score														

Question #1: Identify the correct completions for the following sentence (**Multiple choice**):
Probabilistic graphical models...
[ ] *can be used for inference by means of both exact and approximate reasoning.*
[ ] *cannot deal with a large number of variables.*
[ ] *stand out due to its interpretability.*
[ ] *only work with discrete random variables.*
[ ] *encode the factorization of a joint probability distribution.*
[ ] *are barely useful when the set of (conditional) independencies that the probability distribution of interest fulfills is large.*
[ ] *allow specialists to efficiently deal with partial evidence.* **(____/7)**

Question #2: Form a meaningful sentence by joining each sentence from the first column with its corresponding sentence from the second column (there is a remaining pair of sentences which makes no sense):

(1) Marginalization of a random variable
(2) Factorization of a joint probability distribution
(3) Inference with Bayesian networks
(4) Product of factors
(5) Many BNs of specific purpose
(6) Frequentist statistics
(7) Bayesian statistics

(A) have been developed for the task of classification.
(B) multiplies the values for all the possible instantiations fixed the value of intersecting variables.
(C) assume that a model parameter is an unknown specific value.
(D) sums out over the distribution of the rest of variables.
(E) consider model parameters as random variables.
(F) is the product of the likelihood times the prior distribution.
(G) leads to a product of conditional probability distributions.

1	2	3	4	5	6	7
			B			

(____/6)

Question #3: Find the corresponding sentence: (**Single choice**)
Laplace smoothing...
[ ] *is necessary to do inference with small probability values.*
[ ] *adds an extra count for each instance matching the considered values.*
[ ] *is used to compensate the counts of highly frequent instances.*
[ ] *prevents zero probability values for previously unseen instantiations.*

The use of the logarithm function...
[ ] *is always advisable to be able to use the "sum" operator, which is simpler than the "product".*
[ ] *allows for dealing with large products of probability terms which might not fit into memory.*
[ ] *should only be used in extreme cases as its use involves loss of precision in the calculations.*
[ ] *can safely be used as a substitute of a product when the probability terms are strictly smaller than 1.*

In inference, the evidence spreads throughout the model...
[ ] *to build an update probability distribution over a set of variables of interest (observed variables).*
[ ] *to answer two types of queries: the most-probable assignment (MAP) and the marginal queries.*
[ ] *to obtain the most-probable assignment to the non-observed variables.*
[ ] *from a training dataset iid sampled from the probability distribution of interest.*

Gibbs sampling...

[ ] *is a specific MCMC process where neighborhoods are randomly chosen.*

[ ] *considers the factors of all variables in the Markov blanket when computing the distribution to sample from.*

[ ] *is a specific MCMC process where transitions are determined by complete instantiations of the model.*

[ ] *allows for considering only the affected factors when computing the distribution to sample from.*

The EM algorithm...

[ ] *is guaranteed to reach a global maximum of the likelihood function.*

[ ] *is usually run multiple times using different E-steps to try to find the global maximum.*

[ ] *iteratively alternates two steps: expectation (obtaining model parameters) and maximization (finding the best possible instantiation).*

[ ] *iteratively alternates two steps: expectation (obtaining the probabilistic assignment for missing data) and maximization (obtaining model parameters).*

(____/5)

Question #4: Basic probability.

Calculate the probability $p(A = +|B = +)$ given the following marginal and conditional probability distributions,

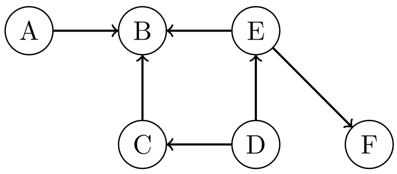
A	$p(A)$
-	0.40
+	0.60

B	A	$p(B\|A)$
-	-	0.50
+	-	0.50
-	+	0.70
+	+	0.30

(____/2)

Question #5: Independencies in Bayesian networks.

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



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

[ ] $B \perp\!\!\!\perp F|E$

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

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

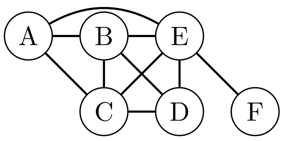
[ ] $C \perp\!\!\!\perp E|D$

[ ] $A \perp\!\!\!\perp F|E$

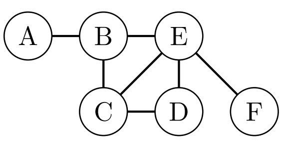
(____/6)

Question #6: Variable Elimination.

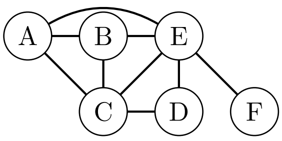
Considering the previous structure G, which of the following structures is the induced graph resulting from following the ordering (A, C, B, D, E, F)? (**Single choice**)



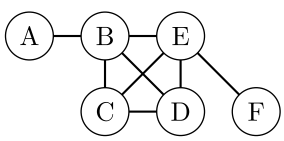
[ ] G_1



[ ] G_2



[ ] G_3

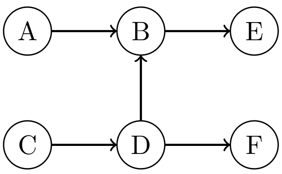


[ ] G_4

(____/1)

Question #7: Factorization in Bayesian networks (I-map).

Consider the following 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**)



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

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

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

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

(____/4)

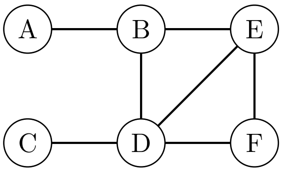
Question #8: Number of parameters in Bayesian networks.

Given the network of the previous exercise, assume that variables A, B and C are binary, variables D and E have four possible values each, and variable F has six possible values. How many parameters are required to encode the probability distribution? Justify your answer.

(____/2)

Question #9: Independencies in Markov networks.

Consider the following structure, H. Write an independence statement satisfied by this Markov network which is not satisfied by the Bayesian network from the previous exercise.



(____/1)

Question #10: Factorization in Markov networks.

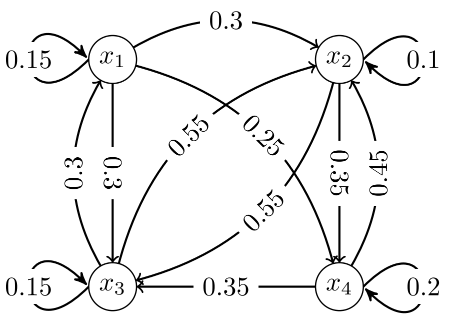
Consider the previous structure, H, and the set of probability distributions, P, which factorize as displayed. Which distribution P is guaranteed to factorize according to H? (**Multiple choice**)

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

(____/4)

Question #11: Stationary distributions.

By definition, which of the following properties must be satisfied by a stationary distribution π for this simple Markov chain? (**Multiple choice**)

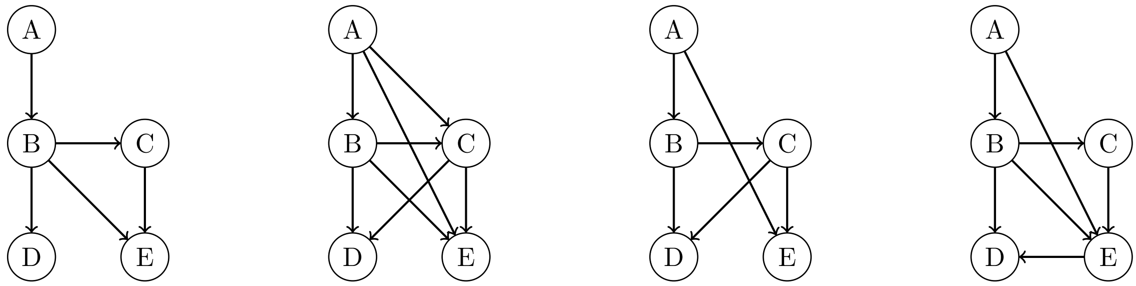


[ ] $\pi(x_1) + \pi(x_2) + \pi(x_3) = 1$
[ ] $\pi(x_1) = 0.15\pi(x_1) + 0.3\pi(x_3)$
[ ] $\pi(x_4) = 0.25\pi(x_1) + 0.35\pi(x_2) + 0.2\pi(x_4)$
[ ] $\pi(x_1) = 0.15\pi(x_1) + 0.3\pi(x_2) + 0.3\pi(x_3) + 0.25\pi(x_4)$
[ ] $\pi(x_4) = 0.45\pi(x_2) + 0.35\pi(x_3) + 0.2\pi(x_4)$

(____/5)

Question #12: Structural learning: Likelihood score.

Given these four Bayesian network structures, which of the following statements about the likelihood scores of the different graphs is true? (**Multiple choice**)



G_1 G_2 G_3 G_4

[ ] $Score_L(G_1 : D) \geq Score_L(G_3 : D)$ for every dataset D.
[ ] $Score_L(G_2 : D) \geq Score_L(G_3 : D)$ for every dataset D.
[ ] $Score_L(G_2 : D) \geq Score_L(G_4 : D)$ for every dataset D.
[ ] $Score_L(G_4 : D) \geq Score_L(G_1 : D)$ for every dataset D. **(____/4)**

Question #13: Parametric learning.

a) Estimate the CPD of $P(Cold|Fever, Headache)$ from the following dataset with 30 instances.

b) Using Bayesian statistics and assuming a prior distribution $Beta(2, 2)$ for $P(Cold|Fever = yes, Headache = no)$, which is the posterior Beta distribution after observing the following dataset?

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

(____/4)