CS536 Winter 2016 Mid-term Exam

Instructions: 15 points total, 4 questions, 50 minutes, no notes. Put your name at the top.

Show your work and final result. Answers only require an explanation if explicitly stated — I am looking for succinct answers; verbose answers will not receive full credit. Provide answers in the space provided; use space at the end if absolutely necessary (clearly label the question #).

1. (a) [1 pt] For a set of continuous random variables \( V \), a set of query variables \( Q \subseteq V \) and a set of evidence variables \( E \subseteq V \) where \( Q \cap E = \emptyset \), show how the joint distribution \( P(V) \) is key to answering all probabilistic queries \( P(Q|E) \), i.e., write the expression for \( P(Q|E) \) in terms of \( P(V) \):

\[
P(Q|E) = \frac{P(Q,E)}{P(E)} = \frac{\sum_{V \setminus E} P(V)}{\sum_{V \setminus E} P(V)}
\]

(b) [1 pt] Assume that I give you a positive function \( F(V) \) and define \( P(V) = \frac{1}{Z} F(V) \) where \( Z \) is a normalizing constant. Briefly show why we do not need to know the value of \( Z \) in order to answer the query \( P(Q|E) \):

\[
P(Q|E) = \frac{\frac{1}{Z} F(V)}{\sum_{V \setminus E} \frac{1}{Z} F(V)}
\]

(c) A graphical model is a compact representation of a joint probability distribution. Consider a naïve approach of representing a joint distribution in a single fully enumerated table of all possible joint variable assignments and their associated probabilities. Relative to this naïve approach, explain a practical motivation for graphical models in terms of each of the following:

(i) [1 pt] Representation space:

- Potential
- Exponential reduction in space

(ii) [1 pt] Computational complexity of inference:

- Potential
- Exponential reduction in time complexity of inference

(iii) [1 pt] Sample complexity of learning:

- Potential
- Exponential reduction in number of parameters to learn
- Reduces sample size needed to learn well (low variance parameter estimates)
2. All questions reference the Bayes net graphical model shown on this page.

(a) [1 pt] Write down the factorized form of the joint distribution for the Bayes net:

\[ P(A, \ldots, H) = P(B \mid A, F) \cdot P(H \mid F) \cdot P(F \mid E) \cdot P(E \mid C, D) \cdot P(B \mid A) \cdot P(C \mid A) \cdot P(A) \cdot P(D) \]

(b) [2 pts] Answer the following conditional independence queries (circle the answer):

a. B ⊥ C? Yes / No
b. B ⊥ C | A? Yes / No

c. B ⊥ C | A, G? Yes / No

d. A ⊥ D | H? Yes / No

(c) Assume all variables are discrete. Currently variable G has two parents (B and F) and no children. Answer the following:

(i) [1 pt] If variable G had 100 additional parents (with no parents themselves), could we use a tabular representation for all factors in the Bayes net? If yes, explain why. If not, provide one possible alternative to a tabular representation and specify where it should be used.

No, exponential soon (≈ 2^100) in tabular representation.
Use noisy-or (independent cases)

(ii) [1 pt] If variable G instead had 100 children (and all of these children had no children themselves), could we use a tabular representation for all factors in the Bayes net? If yes, explain why. If not, provide one possible alternative and specify where it should be used.

Yes, all children tests independent if \( P(\text{child} \mid G) \)
so each \( P(\text{child} \mid G) \) is a small table.
3. (a) [3 pts] Compute the MPE (most probable explanation / assignment to all variables) from the following distributions (assuming A is independent of C given B) using variable elimination (show your work including all intermediate factors produced and show all numerical results as fractions):

\[
P(A|B) = \begin{array}{ccc}
A & B & Pr \\
T & T & .25 \\
T & F & .5 \\
F & T & .75 \\
F & F & .5 \\
\end{array}
\]

\[
P(B|C) = \begin{array}{ccc}
B & C & Pr \\
T & T & .25 \\
T & F & .5 \\
F & T & .75 \\
F & F & .5 \\
\end{array}
\]

\[
P(C) = \begin{array}{c}
C & Pr \\
T & .25 \\
F & .75 \\
\end{array}
\]

\[
\max \max \max P(A|B) \cdot P(B|C) \cdot P(C)
\]

\[
\max \max P(A|B|C) \cdot \max A, B
\]

\[
\max \max P(A|B) \cdot P(B|C) \cdot P(C)
\]
(b) [1 pt] If instead in 3(a), A, B, and C were continuous variables and \( P(A|B), P(B|C), \) and \( P(C) \) were arbitrary continuous distributions, would it be easy to compute the MPE using variable elimination as it was above for the discrete variable case? Why or why not?

\[
\begin{align*}
\text{No. max } F(x) \text{ is optimization which in } x \in \mathbb{R} \\
\text{general is intractable.}
\end{align*}
\]

4. Assume we are given \( m \) binary variables \( A_1, \ldots, A_m \) and \( n \) binary variables \( B_1, \ldots, B_n \). In the tabular representation of \( P(A_1) \), there is only one free parameter we can independently set since once we set \( P(A_1 = \text{true}) = p \) (for some constant \( p \)), we can determine \( P(A_1 = \text{false}) = 1 - p \).

[1 pt] How many free parameters can we independently set in the tabular representation of \( P(A_1, \ldots, A_m | B_1, \ldots, B_n) \)? Briefly explain your reasoning.

\[
2^m \text{ possible contexts} \\
(2^m - 1) \text{ free parameters per context (last is known: sum sums to 1)} \\
2^m (2^m - 1) \text{ free parameters}
\]