Special Topics in AI (CS 7180)
Homework: Probabilistic Graphical Models

Instructor: Ehsan Elhamifar

Due Date: March 31, 2017, 1:35pm

1. Independencies in Bayesian Networks. Consider the model shown in Figure 1. Indicate whether the following independence statements are true or false according to this model. Provide a very brief justification of your answer (no more than 1 sentence).

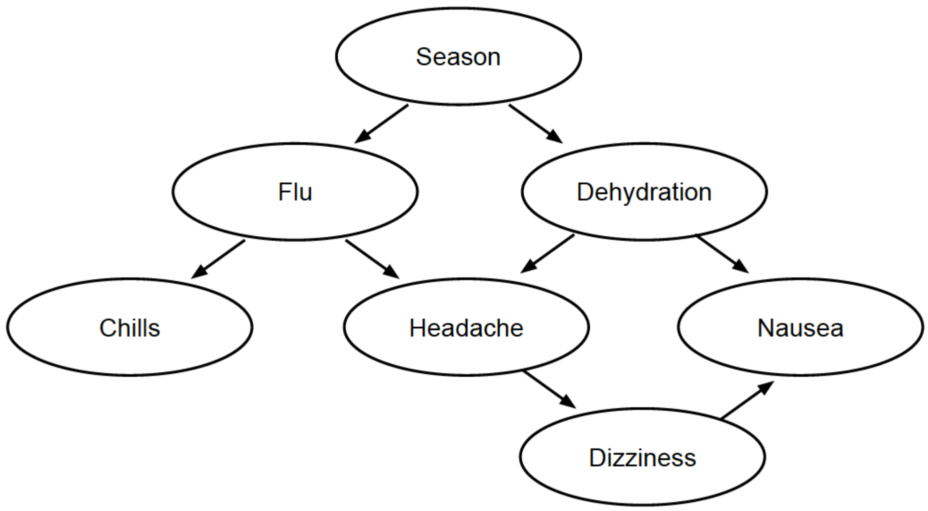


Figure 1: A Bayesian network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

1. Season $\perp$ Chills
2. Season $\perp$ Chills | Flu
3. Season $\perp$ Headache | Flu
4. Season $\perp$ Headache | Flu, Dehydration
5. Season $\perp$ Nausea | Dehydration
6. Season $\perp$ Nausea | Dehydration, Headache
7. Flu $\perp$ Dehydration
8. Flu $\perp$ Dehydration | Season, Headache
9. Flu $\perp$ Dehydration | Season

10. Flu $\perp$ Dehydration | Season, Nausea
11. Chills $\perp$ Nausea
12. Chills $\perp$ Nausea | Headache

2. Factorized Joint Distributions. a) Using the directed model shown in Figure 1, write down the factorized form of the joint distribution over all of the variables, $P(S, F, D, C, H, N, Z)$. b) Using the undirected model shown in Figure 2, write down the factorized form of the joint distribution over all of the variables, assuming the model is parameterized by one factor over each node and one over each edge in the graph.

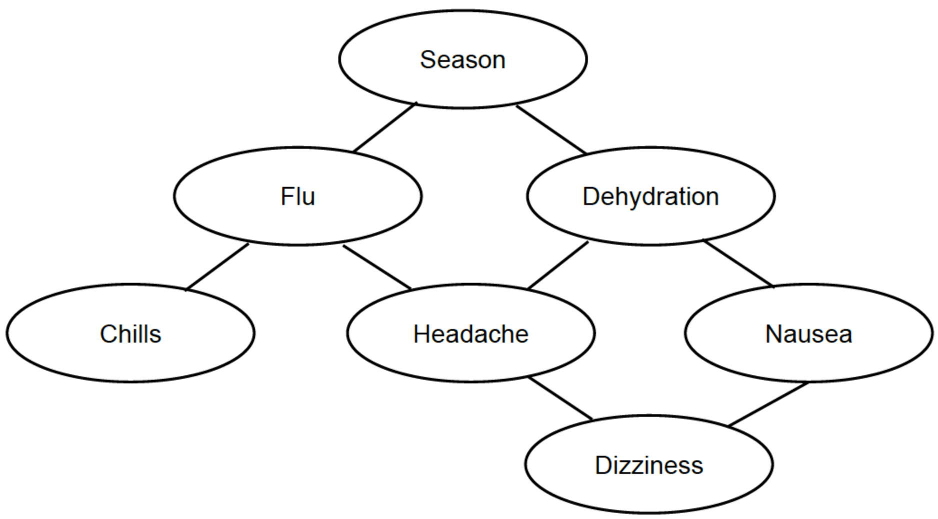


Figure 2: A Markov network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

3. Evaluating Probability Queries. Assume you are given the conditional probability tables listed in Figure 3 for the model shown in Figure 1. Evaluate each of the probabilities queries listed below, and show your calculations.

1. What is the probability that you have the flu, when no prior information is known?
2. What is the probability that you have the flu, given that it is winter?
3. What is the probability that you have the flu, given that it is winter and that you have a headache?
4. What is the probability that you have the flu, given that it is winter, you have a headache, and you know that you are dehydrated?
5. Does knowing you are dehydrated increase or decrease your likelihood of having the flu? Intuitively, does this make sense?

$P(S = \text{winter})$	$P(S = \text{summer})$
0.5	0.5

	$P(F = \text{true} \mid S)$	$P(F = \text{false} \mid S)$
$S = \text{winter}$	0.4	0.6
$S = \text{summer}$	0.1	0.9

	$P(D = \text{true} \mid S)$	$P(D = \text{false} \mid S)$
$S = \text{winter}$	0.1	0.9
$S = \text{summer}$	0.3	0.7

	$P(C = \text{true} \mid F)$	$P(C = \text{false} \mid F)$
$F = \text{true}$	0.8	0.2
$F = \text{false}$	0.1	0.9

	$P(H = \text{true} \mid F, D)$	$P(H = \text{false} \mid F, D)$
$F = \text{true},\ D = \text{true}$	0.9	0.1
$F = \text{true},\ D = \text{false}$	0.8	0.2
$F = \text{false},\ D = \text{true}$	0.8	0.2
$F = \text{false},\ D = \text{false}$	0.3	0.7

	$P(Z = \text{true} \mid H)$	$P(Z = \text{false} \mid H)$
$H = \text{true}$	0.8	0.2
$H = \text{false}$	0.2	0.8

	$P(N = \text{true} \mid D, Z)$	$P(N = \text{false} \mid D, Z)$
$D = \text{true},\ Z = \text{true}$	0.9	0.1
$D = \text{true},\ Z = \text{false}$	0.8	0.2
$D = \text{false},\ Z = \text{true}$	0.6	0.4
$D = \text{false},\ Z = \text{false}$	0.2	0.8

Figure 3: Conditional probability tables for the Bayesian network shown in Figure 1.

4. Bayesian Networks vs. Markov Networks. Now consider the undirected model shown in Figure 2.

1. Are there any differences between the set of marginal independencies encoded by the directed and undirected versions of this model? If not, state the full set of marginal independencies encoded by both models. If so, give one example of a difference.

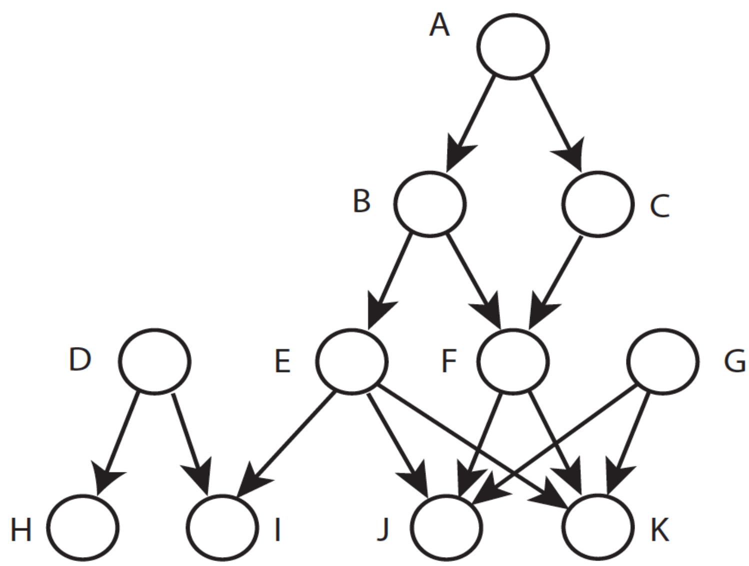
2. Are there any differences between the set of conditional independencies encoded by the directed and undirected versions of this model? If so, give one example of a difference.

5. Ising Model. Consider an $n \times n$ lattice (a two-dimensional grid in which each interior node is linked to its four nearest neighbors). What is the best elimination order for such a lattice? What is the running time of the elimination algorithm for such a best order?

6. Minimal Conditional Independence. For a given variable X_i in a graphical model, what is the minimal set of nodes that renders X_i conditionally independent of all of the other variables? That is, what is the smallest set C such that $X_i \perp X_{V \setminus i \cup C} | X_C$? (Note that V is the set of all nodes in the graph, so that $V \setminus i \cup C$ is the set of all nodes in the graph excluding i and C). (a) Do the problem for an undirected graph. (b) Do the problem for a directed graph.

7. Computing Marginals of Pairs of Unlinked Nodes. Consider an undirected graph, $G = (V, E)$, and consider a pair of nodes X_i and X_j that are not linked in G; i.e., for which $(i, j) \notin E$. How can you use the junction tree algorithm to compute the marginal $p(x_i, x_j)$?

8. Junction Tree. Show how to moralize the following graph, and construct a junction tree. What is the worst-case complexity of computing marginal probabilities with your junction tree?



9. Exponential Family. A probability distribution in the exponential family takes the following general form: $p(x|\eta) = h(x) \exp \eta^T T(x) ? A(\eta)$ for a parameter vector η, often referred to as the natural parameter, and for given functions T, A, and h.

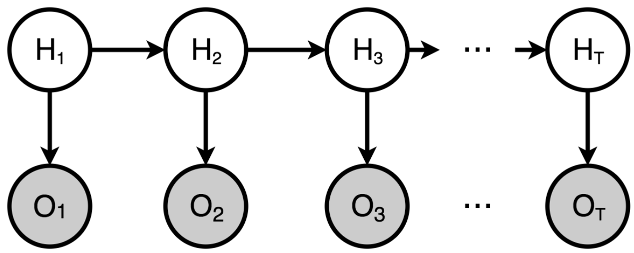
(a) Show that the following distributions are in the exponential family, exhibiting the T, A and h functions in each case.

1. Poiss(λ) – Poisson with parameter λ
2. $\mathcal{N}(\mu, I)$ – multivariate Gaussian with mean vector μ and identity covariance matrix.
3. Mult(θ) – multinomial with parameter vector $\theta = (1, 2, \ldots, K)$. Use the fact that $\theta_K = 1 - \sum_{k=1}^{K} \theta_k$ and express the distribution using a $(K-1)$-dimensional parameter η.

(b) The function $A(\eta)$ turns out to have moment-generating properties. In particular, show the following: $\Delta_\eta A = E[T(X)]$.

(c) Demonstrate that the relationship in (b) holds for the three examples in part (a).

10. Inference in HMMs. In a Hidden Markov Model (HMM), we have a Markov Chain on states $(H_0, \ldots, H_T)$ which are often not observed. Instead, at each time instant t, we observe some function of the state H_t, denoted by O_t.



(a) For each of the following statements, verify if it is true or false.

1. $H_{t+i} \perp H_{t-j}, \ \forall i, \ j \ \geq 1$
2. $H_{t+i} \perp H_{t-j} \mid H_t, \ \forall i, \ j \ \geq 1$
3. $O_{t+i} \perp O_{t-j}, \ \forall i, \ j \ \geq 1$
4. $O_{t+i} \perp O_{t-j} \mid H_t, \ \forall i, \ j \ \geq 1$

(b) Using the chain rule of Bayesian Networks, write down the joint distribution of $p(H_{0:T}, O_{0:T})$, where $H_{0:T} \triangleq H_0, \ldots, H_T$ (and similarly $O_{0:T}$).

(c) Build a Cluster Graph (CG) with its cluster zero consisting of H_0, cluster t consisting of $\{H_{t-1}, H_t\}$ for $t = 1, \ldots, T$. Connect each cluster t to cluster $t-1$ and $t+1$.

1. Draw the corresponding CG with clusters and sepsets and assign initial potentials to each cluster using $p(H_0)$, $\{p(H_t|H_{t-1})\}_{t=1}^T$ and $\{p(O_t|H_t)\}_{t=0}^T$.
2. Using Belief Propagation Algorithm, write down messages $\delta_{t\to t+1}$ and $\delta_{t+1\to t}$ for two cases of $t = 0$ and $t \geq 0$. Which variable(s) are $\delta_{t\to t+1}$ and $\delta_{t+1\to t}$ a function of?
3. Compute $p(O_{0:T}), \ p(H_t|O_{0:T})$ and $p(H_t, H_{t+1}|O_{0:T})$ using messages. The resulting method is also called the $\alpha - \beta$ algorithm in the literature.