Quiz 1

Date: Monday, October 17, 2016

Name: ________________________________

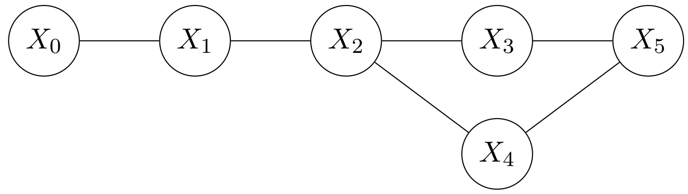
Andrew ID: ________________________________

Department: ________________________________

Guidelines:

1. **PLEASE DO NOT TURN THIS PAGE UNTIL INSTRUCTED.**
2. Write your name, Andrew ID, and department in the spaces provided above.
3. You have **sixty (60)** minutes for this exam.
4. This exam has **seven (7)** pages on seven (7) sheets of paper, including this one.
5. This exam has a total of 50 possible points. The number of points allocated to each question is indicated next to each question.
6. This exam is open notes. You may use any materials such as cheat sheets, class notes, etc. No electronic devices are permitted.
7. The questions vary in difficulty. The points allocated to a question do not entirely reflect its difficulty. Do not spend too much time on one question.
8. Questions only appear on one side of each sheet of paper. You may use any blank space for your answer or scratch work, but please clearly indicate your answers.

1. **[10 points]** Consider the undirected graphical model shown below:



(a) Which of the following statements are always true? No justification is needed.

i. **(2 points)** $H(X_0|X_1) \leq H(X_0|X_2)$

ii. **(2 points)** $I(X_0; X_3) \leq I(X_0; X_5)$

iii. **(2 points)** $I(X_2; X_5|X_4) \leq I(X_2; X_3|X_4)$

(b) **(4 points)** Suppose we observe n IID samples from the joint distribution of $(X_0, \ldots, X_5)$, and use the Chow-Liu algorithm with a consistent mutual information estimator. Explain why we never recover the above graph structure, even as $n \to \infty$.

2. **[12 points]** Let X_i for $i \in [d] = \{1, \ldots, d\}$ be independent random variables. Show that

(a) **(3 points)** Show that $I(X_i; X_i + X_j) = H(X_i + X_j) - H(X_j)$.

(b) **(4 points)** Show that $I(X_i; X_i + X_j) \geq I(X_i; X_i + X_j + X_k)$.

(c) **(5 points)** Define $f : 2^{[d]} \to \mathbb{R}$ by

$$f(S) := H\left(\sum_{i \in S} X_i\right), \quad \forall S \subseteq [d].$$

Show that f is submodular.

3. **[8 points]** Suppose you flip a coin independently n times and observe cn heads and $(1-c)n$ tails. Explain how to use the MDL principle to choose the best model amongst $M = \cup_\ell M_\ell$, where

$$M_\ell : \text{ The probability the coin lands heads is } z2^{-\ell} \text{ for some integer } z \in [0, 2^\ell).$$

Write the MDL rule in terms of n, c, z and ℓ only.

4. **[10 points]** Suppose we already have an estimate $\widehat{p}$ for some probability density p on $\mathcal{X}$. Using n new IID samples $X_1, \ldots, X_n \sim p$, we want to estimate the squared L_2-norm

$$\|p\|_2^2 = \int_{\mathcal{X}} p^2(x)\, dx = \mathbb{E}_{X \sim p}\left[p(X)\right].$$

of p. Show that the first-order von Mises estimator is identical to the re-substitution estimator:

$$\frac{1}{n} \sum_{i=1}^{n} \widehat{p}(x),$$

5. **[10 points]** Consider a set of k variables $X_1, \ldots, X_k$, and suppose we know the pairwise distributions $p(X_i, X_{i+1})$, for $i \in \{1, \ldots, k-1\}$, of consecutive pairs. Show that the MaxEnt joint distribution $p(X_1, \ldots, X_k)$ is a first-order Markov chain (i.e., for any $i_1 < i_2 < i_3$ in $\{1, \ldots, k\}$, X_{i_1} and X_{i_3} are conditionally independent given X_{i_2}). *(Hint: Write the joint distribution $H(X_1, \ldots, X_k)$ in terms of $\sum_{i=1}^{k} H(X_i|X_{i-1})$+another term.)*

6. **[Optional - no credit]** If you found the quiz too easy, prove the following for problem 1. Assume all the edge weights are distinct. Argue that, as $n \to \infty$, we always recover the edge $X_0 - X_1$. *(Hint: Argue by means of contradiction.)*

Please do not mark below this line.

Problem	Max	Points
Q1	10	
Q2	12	
Q3	8	
Q4	10	
Q5	10	
Total	50	