1. (Modified version of Exercise 3.1 in the textbook) Suppose you have three Boolean random variables $X_1, X_2, X_3$. Write out a discrete joint probability distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in I(P)$, but we also have that $((X_1, X_2) \perp X_3) \notin I(P)$. This should be a table with $2^3$ rows (one for each combination of values for the three random variables). Explain your solution. [4 points]

2. (Modified version of 3.6) Consider a set of random variables $X_1, X_2, \ldots, X_n$ where each $X_i$ has $|Val(X_i)| = l$ i.e. $X_i$ takes on $l$ different values.
   a) Assume that we have a Bayesian network over $X_1, X_2, \ldots, X_n$ such that each node has at most $k$ parents. What is a simple upper bound on the number of independent parameters in the Bayesian network? How many independent parameters are in the full joint distribution over $X_1, X_2, \ldots, X_n$? [2 points]

   b) Now, assume that each variable $X_i$ has the parents $X_1, X_2, \ldots, X_{i-1}$. How many independent parameters are there in the Bayesian network? What can you conclude about the expressive power of this type of network? [4 points]

   c) Now, consider a naïve Bayes model where $X_1, X_2, \ldots, X_n$ are evidence variables (i.e. observed features), and we have an additional class variable $C$ which has $k$ possible values $c_1, \ldots, c_k$. How many independent parameters are required to specify the naïve Bayes model? How many independent parameters are required for an explicit representation of this joint distribution? [4 points]

3. Answer true or false to the following conditional independence statements using the graph below. For partial credit, show the paths that are blocked or not blocked. [9 points]
   a) $A \perp D \mid \{C, F\}$
   b) $B \perp F \mid A$
   c) $A \perp C \mid \{B, E, F, G\}$
4. (Exercise 3.7 in the textbook) Show how you could efficiently compute the distribution over a variable $X_i$ given some assignment to all the other variables in the network: $P(X_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$. Your procedure should not require the construction of the entire joint distribution $P(X_1, ..., X_n)$. Specify the computational complexity of your procedure using Big-O notation. [8 points]

5. (Exercise 3.10) Prove that the global independencies, derived from d-separation, imply the local independencies (i.e. a node is conditionally independent of its non-descendants given its parents). In other words, prove that a node is d-separated from its non-descendants given its parents. [9 points]