Part C / MSc

Course title: Graphical Models SC6/SM9
Lecturer: Robin Evans

Date of this version: April 25, 2018

ABD: EveryShipout initializing macros

Do not turn this page until you are told that you may do so

1. Let $\mathcal{G} = (V, E)$ be an undirected graph, and $p(x_V)$ a probability density for random variables $(X_v)_{v \in V}$.

A *path* in $\mathcal{G}$ is a sequence of distinct vertices $v_0, \ldots, v_l$ such that $\{v_{i-1}, v_i\} \in E$ for each $i = 1, \ldots, l$. If π is a path then we say that π' is a *subpath* of π if π' is a path in $\mathcal{G}$ and a subsequence of π.

(a) [2 marks] Explain what it means to say that p satisfies the global Markov property with respect to $\mathcal{G}$.

(b) [3 marks] What does it mean to say that (A, S, B) is a *proper decomposition* of $\mathcal{G}$? ‘

Suppose now that $\mathcal{G}$ is a graph with a proper decomposition (A, S, B).

(c) [5 marks] Show that any path π from $v \in A \cup S$ to $w \in A \cup S$ is either contained in $A \cup S$, or there is a subpath π' of π, also from v to w, such that π is contained in $A \cup S$.

(d) [5 marks] Hence show that if $p(x_V)$ is globally Markov with respect to $\mathcal{G}$, then $p(x_{A \cup S})$ is globally Markov with respect to the induced subgraph $\mathcal{G}_{A \cup S}$.

(e) [3 marks] What does it mean to say that $\mathcal{G}$ is decomposable?

(f) [7 marks] Let $\mathcal{G}$ be a decomposable graph. Show that if p is globally Markov with respect to $\mathcal{G}$ then

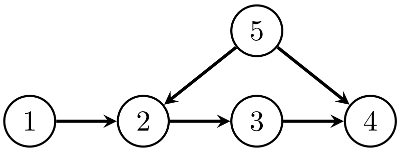
$$p(x_V) = \prod_{i=1}^{k} \frac{p(x_{C_i})}{p(x_{S_i})}$$

for suitable sets $C_1, \ldots, C_k$ and $S_1, \ldots, S_k$ that you should define.

[You may assume standard results about decomposable graphs, provided they are stated clearly.]

2. (a) [4 marks] Let $\mathcal{G}$ be a directed acyclic graph with vertices V. Explain what it means for a distribution p to *factorize* with respect to $\mathcal{G}$, defining any graphical terminology you use.

(b) [6 marks] Now consider the graph $\mathcal{G}$ in the figure below, and suppose that the distribution $p(x_1, x_2, x_3, x_4, x_5)$ factorizes with respect to $\mathcal{G}$.



For each of the conditional independence statements below, explain whether or not it is implied by the graph. You may assume equivalence of any suitable Markov properties, provided that you state clearly which one you are using and how.

(i) $X_1 \perp\!\!\!\perp X_5$;

(ii) $X_1 \perp\!\!\!\perp X_4 \mid X_3$;

(iii) $X_3 \perp\!\!\!\perp X_1, X_5 \mid X_2$.

(c) [7 marks] Define the interventional distribution $p(x_2, x_4, x_5 \mid do(x_1, x_3))$ with respect to the graph $\mathcal{G}$. Show that your definition is equivalent to

$$p(x_2, x_4, x_5 \mid do(x_1, x_3)) = \frac{p(x_1, x_2, x_3, x_4, x_5)}{p(x_1) \cdot p(x_3 \mid x_2)}.$$

(d) [8 marks] Deduce that

$$p(x_4 \mid do(x_1, x_3)) = \sum_{x_2} p(x_2 \mid x_1) \cdot p(x_4 \mid x_1, x_2, x_3) \tag{1}$$

and, by considering a different but equivalent expression for $p(x_4 \mid do(x_1, x_3))$, show that the quantity in (1) does not depend upon x_1.

3. Let $\boldsymbol{X} = (X_1, \ldots, X_p)^T \sim N_p(0, \Sigma)$ be a vector with a multivariate Gaussian distribution, where the $p \times p$ covariance matrix Σ is positive definite.

(a) [4 marks] Define the *concentration matrix* K, and write down the density function of $\boldsymbol{X}$ in terms of K.

(b) [5 marks] Derive the conditional distribution of $X_1 \mid X_2 = x_2, \ldots, X_p = x_p$.

(c) [3 marks] Deduce that $X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$ if and only if $k_{ij} = 0$.

(d) [6 marks] Define the *Gibbs sampling algorithm*, and explain how you could use it to obtain samples from this distribution.

(e) [2 marks] Suppose that K is a given matrix for some large number p, and that Σ is not explicitly known. Given one advantage and one disadvantage of obtaining samples using Gibbs.

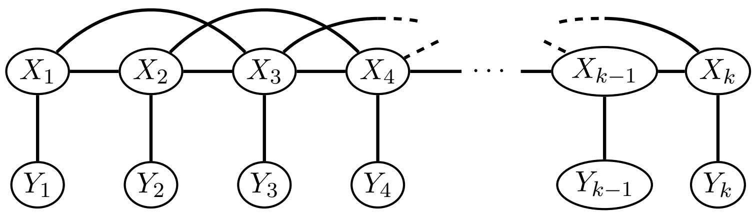
4. (a) [5 marks] What does it mean to say that two potentials $\psi_C(x_C)$ and $\psi_D(x_D)$ with $S = C \cap D \neq \emptyset$ are *consistent*? Explain the process of *passing a message* from a ψ_C to ψ_D.

(b) [3 marks] Show that

(i) after passing a message from ψ_C to ψ_D, the updated potentials ψ_C and ψ_S will be consistent;

(ii) if ψ_D and ψ_S were consistent before message passing, then they remain so afterwards.

Now consider the following variation on a hidden Markov model, in which X_i is adjacent to $X_{i-2}, X_{i-1}, X_{i+1}, X_{i+2}$ as well as Y_i:



(c) [3 marks] Draw a junction tree suitable for inference in this model.

(d) [7 marks] Give an algorithm that will make the potentials associated with the junction tree consistent, whilst preserving a suitable function of the potentials. Explain, with reference to part (b), how the algorithm does this.

(e) [2 marks] What is the advantage of obtaining consistency?

(f) [5 marks] Suppose that your junction tree is consistent. How would you calculate the conditional distribution of $X_1, \ldots, X_k$ given $\{Y_1 = y_1, \ldots, Y_k = y_k\}$?

5. Let $\mathcal{G}$ be a directed acyclic graph with vertex set V, and let p be a probability density over random variables X_V.

(a) [4 marks] What does it mean to say that p satisfies the *local Markov property* for $\mathcal{G}$? Be sure to define any graphical concepts and notation that you use.

(b) [4 marks] Let $(X, Y, Z)^T$ have a joint Gaussian distribution such that $X \perp\!\!\!\perp Z \mid Y$. Show that

$$\operatorname{Cov}(X, Z) = \frac{\operatorname{Cov}(X, Y) \operatorname{Cov}(Y, Z)}{\operatorname{Var} Y}.$$

[You may use without proof the fact that, if $(X_1, \dots, X_k)^T$ has a multivariate Gaussian distribution, then $X_k = \sum_{i=1}^{k-1} \beta_i X_i + \varepsilon_k$ for some $\varepsilon_k \perp\!\!\!\perp X_1, \dots, X_{k-1}$, and constants $\beta_1, \dots, \beta_{k-1}$.]

(c) [7 marks] Let $\mathcal{G}$ contain a path π: $1 \to 2 \to \cdots \to k$ for $k \geqslant 3$, and let $C \subseteq V$ be a set not containing any of the vertices on π.

By induction or otherwise, show that there exists a distribution p satisfying the local Markov property for $\mathcal{G}$ such that $X_1 \not\!\perp\!\!\!\perp X_k \mid X_C \, [p]$.

[You **may not** make use of other Markov properties, unless you first show their equivalence to the local Markov property.]

Suppose that p satisfies the local Markov property for $\mathcal{G}$.

(d) [2 marks] Let $i, j \in V$ not be joined by an edge in $\mathcal{G}$. Give a necessary and sufficient condition for the independence $X_i \perp\!\!\!\perp X_j \mid X_{\mathrm{pa}_{\mathcal{G}}(i)}$ to be implied by the local Markov property for $\mathcal{G}$. Prove your assertion.

(e) [5 marks] Write an explanation of how the 'skeleton' phase of the PC algorithm works, with access to an oracle independence test.

(f) [3 marks] Prove that the skeleton phase will return the correct skeleton.