Probabilistic Modelling and Reasoning, Tutorial Answer Sheet 1 (for Week 3)

School of Informatics, University of Edinburgh

Instructor: Amos Storkey
Prob sheet ID: 143025.1.1

Please answer all the questions. We expect to get through Q1, Q2, and probably Q3 in the tutorial.

1. Write
$$P(r, x, y, z) = \frac{1}{Z} P(r|x, z)\Phi_1(x, y)\Phi_2(y, z)$$
where $P(r = 1|x, z) = 0.5 + 0.2(x - z)$, $\Phi_1(x, y) = (x + y)$ and $\Phi_2(y, z) = (y + 1)(z + 1)$ and all r,x,y,z are binary $\{0, 1\}$ variables.

   (a) Compute the value of Z as efficiently as you can (consider using a repeated elimination approach).

   The key to this is to eliminate things in the right order. Now Z must normalise the distribution and so
$$Z = \sum_{r,x,y,z} P(r|x, z)\Phi_1(x, y)\Phi_2(y, z).$$
   We can see this is the right thing to do as then $\sum_{r,x,y,z} P(r, x, y, z) = Z/Z = 1$.
   If we do the sum over r first, this sum is just over the $P(r|x, z)$ term (explain how we can take the other terms out of the sum), and the sum over this term is equal to 1, regardless of the values of x and z as this term is a probability distribution over r. So doing that sum first just leaves
$$P(x, y, z) = \frac{1}{Z}\Phi_1(x, y)\Phi_2(y, z)$$
   Now we can sum over z. This involves summing over the $\Phi_2(y, z)$ term, which results in $\Phi_2^*(y) = \sum_z \Phi_2(y, z) = 3(y + 1) = 3y + 3$. Now we can sum $\Phi_1(x, y)$ over x to get $\Phi_1^*(y) = 1 + 2y$, leaving
$$P(y) = \frac{1}{Z}\Phi_1^*(y)\Phi_2^*(y) = \frac{1}{Z}(1 + 2y)(3y + 3) = \frac{1}{Z}6y^2 + 9y + 3.$$
   As this is a probability, we must have $Z = \sum_y 6y^2 + 9y + 3 = 3 + 18 = 21$. Again in all the above it is important to understand how the sums *distribute* so we can get away with summing over just the mentioned terms. Note that the order we do the computation in affects the ease of that computation.

   (b) Write down all the conditional independence relationships of the form $I(A, B|C)$ that you think might hold (remember we can condition on the null set too, to arrive at unconditional dependence relationships).

If we don't condition on r then we have a simple relationship between x, y and z where y couples the x and the z. If we know y then x and z are independent so $I(x, z|y)$. Conditioning on r results in all the variables being coupled so there are no conditional independence relationships if r is conditioned on.

On the other hand, if we know x and z, that is enough to determine r. Any information from y doesn't help, so $I(r, y|x, z)$.

(c) Draw a minimal[1] undirected and a minimal mixed (directed+unidrected) factor graph for this distribution.

The undirected graph has three factors, denoted $1, 2, 3$ with connections $1 \bullet\!\!-\!\!\circ\ x, z, r$, $2 \bullet\!\!-\!\!\circ\ x, y$ and $3 \bullet\!\!-\!\!\circ\ y, z$. Draw this as a factor graph (factors as square blocks, variables as circles). To create a mixed graph, convert the edge from z to 1 to an arrow pointing at the factor 1. Likewise the edge from x to 1. Finally the edge from 1 to r is converted to an arrow pointing as the variable r. This captures the fact that $P(r|x, z)$ is a conditional distribution.

(d) Test your proposed conditional independence relationships using the factor graph separation rule.

For $I(x, z|y)$, then there are two paths from x to z, one via y which is blocked because y is in the conditioning set, and one via r which is blocked because the factor 1 is 'head to head' along the path, and neither it nor it's descendant r are in the conditioning set. So $I(x, z|y)$. If we had done this with the undirected graph, the path via r would not be blocked. Hence just using the undirected graph loses some of the conditional independence relationships in the directed graph.

For $I(y, r|x, z)$ consider the undirected graph. There are two paths from r to y, one each via x and z. both are blocked as both x and z are in the conditioning set. So $I(y, r|x, z)$. The same applies in the mixed graph case.

2. A multivariate Gaussian distribution can be written as

$$P(\mathbf{x}) = \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

where Σ^{-1} is symmetric.

Draw a minimal factor graph representation for a three dimensional Gaussian distribution. You may find it useful to write out the vector-matrix-vector computation in summation form.

Write $(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \sum_{ij} (x_i - \mu_i)(\boldsymbol{\Sigma}^{-1})_{ij}(x_j - \mu_j)$. There is a factor for each i, j pair with $j \geq i$. The symmetry of $\boldsymbol{\Sigma}^{-1}$ means that the $j < i$ terms are identical to, and can be combined with, the equivalent $j \geq i$ term. If $i = j$ the factor only contains the single variable x_i. Otherwise each factor (i, j) contains the two variables x_i, x_j. This can be drawn as a bipartite factor graph where each factor is a pair of nodes. We could include factors that connect to the singleton nodes, but we could also absorb them into the two-variable factors that already include that variable: the absorbing does not add any additional links so it loses nothing to do it. It is worth clearly working through what the factors will look like after this absorbing process. The final result is a factor graph with three factors: one for each off diagonal term of the $\boldsymbol{\Sigma}^{-1}$ matrix.

[1] Here we use the term "minimal" in a vague way to just to mean a choice of representation that does not produce factors that are bigger (i.e. contain more variables) than they need to be.

3. Consider the following model for images. Each image is split into 8 by 8 pixel regions. Each region i is given a label x_i relating to the type of object that is in that region. Then the 64 pixels y_{ij} $j = 1, 2, \ldots, 64$ in that region i are determined dependent on the label. The probability of an object label in each region i is considered to be directly dependent only on the x_i in neighbouring regions, denoted as the set $\mathcal{N}_i$. Illustrate the form of factor graph you may use for such a model.

   Draw a factor graph with one factor between each x_i and all the y_{ij} for that same i. Then draw another factor above each x_i but also connected to each node in $\mathcal{N}_i$. You can make the links from each x to corresponding y terms (via the factor) directed.

   We can write the distribution out in the form

$$P(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \prod_i P_i(\mathbf{y}_i | x_i) \Psi_i(x_i, \mathbf{x}_{\mathcal{N}_i})$$

   By considering two neighbouring pixels in different regions, and thinking in terms of conditional independence relationships give a criticism of this form of model.

   Suppose y_{18} and y_{21} are neighbouring pixels. Then $I(y_{18}, y_{21} | x_1, x_2)$. This implies knowing the object being represented in each patch makes these pixels independent. But suppose this was the same object. Then there is probably some continuity of texture across the patch boundary and so these are likely to still be dependent, even given we know what object is being represented (note that in fact we have more generally $I(y_{18}, y_{21} | x_1)$).

4. Again, consider the model

$$P(r, x, y, z) = \frac{1}{Z} P(r | x, z) \Phi_1(x, y) \Phi_2(y, z)$$

   where $P(r = 1 | x, z) = 0.5 + 0.2(x - z)$, $\Phi_1(x, y) = (x + y)$ and $\Phi_2(y, z) = (y + 1)(z + 1)$ and all r,x,y,z are binary $\{0, 1\}$ variables.

   (a) Use the elimination algorithm to compute $P(x | z = 1)$.

   (b) EXTRA: Use the elimination algorithm to compute $P(y | r = 0)$.

   Both of these are repetitions of the elimination process we used for computing Z. We have

$$P(x, y, z) = \frac{1}{Z} \Phi_1(x, y) \Phi_2(y, z)$$

   and so we then fix $z = 1$ to get $\Phi_2(y) = 2(y + 1)$, so

$$P(x, y | z) = \frac{1}{Z'} \Phi_1(x, y) \Phi_2(y).$$

   Note that this conditioning results in a new normalisation constant (the distribution is now over new variables). Now summing over y gives us

$$P(x | z) \propto \sum_y (x + y)[2(y + 1)] = 2x + 4(x + 1) = 6x + 4$$

   To find the constant of proportionality we use $Z' = \sum_x (6x + 4) = 10 + 4 = 14$. So $P(x | z) = (6x + 4)/14$.

   (c) **The EXTRA question is for students to work through to convince themselves they are now confident in handling these sorts of questions. Answers will not be provided: try to work on the material until you are confident in your answer to this question.**