Homework 2

Assigned: 03/10/2011.
Due: 14/10/2011.

Exercise 1. Explaining Away

Recall the sprinkler network ("Did it Rain Tonight" lecture slides). An important effect with probabilistic reasoning, which can clearly be demonstrated in human subject studies, but is hard to replicate in non-probabilistic expert systems, is *explaining away*. In a sense, this effect is what makes the semantics of causal (Bayesian) networks a bit tricky at first sight (see Exercise 2), and which contributes to the complicated nonlocal couplings in posterior distributions. If several (unobserved) causes can explain some effect, then evidence of this effect having happened creates a dependency between the causes: each of them becomes more likely than *a priori*. Further evidence of effects that could be caused by some, but not others, will strongly decrease posterior probabilities of the latter. Such causes are "explained away", even though (in fact, precisely *because*) there is no causal link between them and the latter effect.

1. [**3 points**] Recall the variables H (Holmes' grass wet?), R (rain last night?), S (Holmes sprinkler ran?), W (Watson's grass wet?). Compute $P(H = y)$. Compute $P(R = y|H = y)$, $P(S = y|H = y)$, compare them to the prior probabilities. Compute the table $P(R, S|H = y)$, show that R and S are dependent, given $\{H = y\}$.
   Compare $P(R = y|S = y, H = y)$ with $P(R = y|S = y)$. I am confused: "Holmes' sprinkler does not cause rain to fall, even if Holmes finds his grass wet". Clear up this confusion (1–2 sentences).

2. [**4 points**] Compute $P(R = y|H = y, W = y)$, $P(S = y|H = y, W = y)$, compare them to $P(R = y|H = y)$, $P(S = y|H = y)$. Relate your findings to *explaining away*. What is $P(R = y|H = y, W = n)$, $P(S = y|H = y, W = n)$?
   The CPT has $P(W = y|R = n) = 2/10$. Suppose that $P(W = y|R = n) = p \in [0, 1]$ instead (everything else in the CPTs stays the same). What is $P(S = y|H = y, W = y)$ in this case (as function of p)? Explain your finding, its values for $p = 0$ and for $p = 1$.

Exercise 2. Semantics of Bayesian Networks (d-separation)

Recall that the graph part of a graphical model is equivalent (but more convenient to work with) to a list of conditional independence constraints: the conditional independence *semantics* of the graph. Make sure you understand what "constraint" means in this case (and in general in mathematics). A distribution is consistent with the graph if it meets all (conditional independence) constraints implied by the graph. It may meet other constraints (for example, a completely factorized distribution is consistent with any graph), but it must not violate any of them. Think about the graph as a scaffold, a set of preconditions, which message-passing inference algorithms use to distribute the task, independent of the distribution that is fed to them later on (as long as it's consistent).
Conditional independence statements have the form $A \perp B|\mathcal{C}$, A and B variables, $\mathcal{C}$ a set of variables, all disjoint. Literally, conditional independence means that information flow between A and B is blocked by evidence on $\mathcal{C}$ (*i.e.* , knowing the value of $\mathcal{C}$): if you know $\mathcal{C}$, then w.r.t. knowledge of A, it does not help to know B as well (and vice versa). Not surprisingly, this is encoded by *separation* in graphs: $A \perp B|\mathcal{C}$ if there is no path from A

to B that is not blocked by some variable in $\mathcal{C}$ (here, in a directed graph, you can walk along an edge in both directions). Note that $\mathcal{C}$ can be empty, in which case we talk about independence per se (unconditional).

In undirected graphical models (Markov random fields), graph separation is just as expected. But looking at Exercise 1, this cannot work for directed models (Bayesian networks). The causes R and S are independent (*i.e.* , blocked) with no evidence, or also if W alone is observed (which does not lie on a path between them). However, once H gets observed, separating them in the "normal" way, they become *dependent* (explaining away). The reason is the configuration $R \to H \leftarrow S$, turning usual graph separation on its head. But directed model graph separation can still be defined using local rules, as we will explore here.

1. [**3 points**] Given three nodes A, C, B, there are three different graphs with C between A and B: $A \to C \to B$ (1), $A \leftarrow C \to B$ (2), and $A \to C \leftarrow B$ (3). For each of these, consider the two cases "C observed" (a) and "C not observed" (b), and argue whether we should assert the independence of A and B [In order to show that we cannot, give an example of a distribution consistent with the graph, under which A and B are not independent (conditioned on C in case (a), unconditional in case (b)). In order to show we can, show that the definition of (conditional) independence is fulfilled for arbitrary distributions consistent with the graph]. What does that mean in terms of graph separation (C acting on path between A and B)?

2. [**3 points**] You will have noticed that (3) behaves differently from the others, somewhat the opposite. You know why this is the case (explaining away). It is called a *v-node* configuration. A and B are *dependent*, given C. But is that all? Consider $A \to C \leftarrow B$ as part of a larger network, extending downstream of C (see Figure 1, left). Without any evidence, does the interpretation still hold? How about C being unobserved, but evidence in the downstream part? State the the rule precisely when we can assert independence of A and B in this case (no need for complete example distributions, but give sound arguments).
   Hint: For creating counterexamples, deterministic distributions such as $P(X|Y) = \mathrm{I}_{\{X=Y\}}$ are useful. Go back to your argument *why* for (3), not knowing C leaves A, B independent. With further nodes downstream, when exactly can you still use this argument?

3. [**3 points**] Consider the Bayesian network of Figure 1, right. Does it imply the following statements? [If so: argue (why are all paths blocked? If not: state the *shortest* path that is not blocked]

   - $A \perp B | H$
   - $A \perp B | F$
   - $A \perp H | E$

   State the smallest set of nodes which blocks X and Y [argue your point; $\emptyset$ is a set as well]

   - $X = C, Y = D$
   - $X = A, Y = H$
   - $X = A, Y = F$

Exercise 3. DIRECTED AND UNDIRECTED GRAPHICAL MODELS
Directed and undirected graphical models are different formalisms. Not only is one of them much more convenient to use in certain situations than the other, and vice versa, but their implied conditional independence semantics are different.

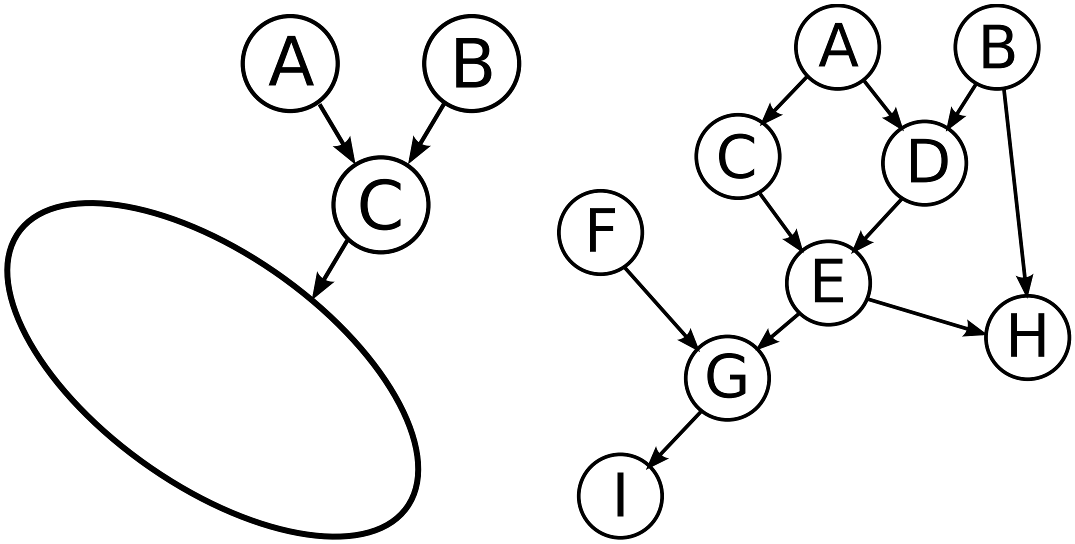

Figure 1: Left: V-node configuration, as part of larger graph. Right: Example Bayesian network

1. **[3 points]** Given a Bayesian network (directed graphical model), state a procedure for converting it into a Markov random field (undirected graphical model). Make sure you understand that *conversion* (in this context) means that:

   - Every distribution consistent with the directed graph has to be consistent with the resulting undirected graph as well. In other words, all conditional independence statements implied by the undirected graph are also implied by the directed graph (not necessarily the other way around).
   - The resulting undirected graph is minimal in fulfilling this requirement. None of its true subgraphs (obtained by removing edges) does the job. More precisely, for every true subgraph, there exists a distribution consistent with the Bayesian network, but not with the undirected subgraph.

   Hint: Remember the Hammersley-Clifford theorem from the lecture, which precisely characterizes the set of distributions consistent with an undirected graph, in terms of potential functions on maximal cliques. For a directed graphical model, how do the potential functions look like?

2. **[1 point]** Apply this procedure to the directed graph $A \to C \leftarrow B$ (v-node configuration, explaining away). State a conditional independence relationship implied by the directed, but not by the undirected graph. This is an example of a directed graph which cannot be exactly represented by an undirected graph.

3. **[2 points]** Consider the undirected model of Figure 2. Prove that no directed model can imply the same conditional independence constraints. State a directed conversion (in the sense of a)).
   Hint: Look for v-nodes.

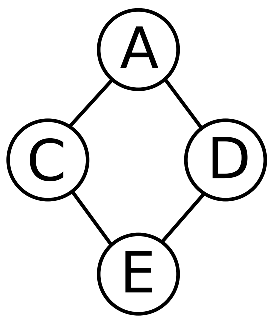

Figure 2: Undirected graphical model which cannot be exactly represented by a directed model.