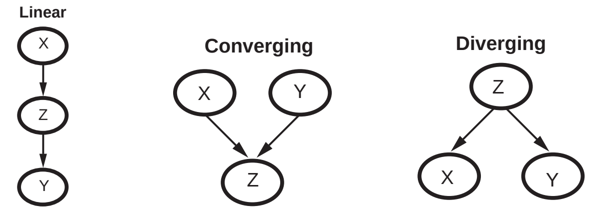
Bayesian Networks — Introductory Remarks II

- Administrative: (i) please post for this week, (ii) this week largely technical set-up for last few weeks (which will involve philosophical controversy and application of network models), (iii) paper topics (let's talk about these soon).
- Evidential Bayesian Networks Overview
  - Conditional Independence Structures and DAGs
  - Evidential Probability Calculations and Traditional PC
- *Causal* Bayesian Networks
  - The Notion of Intervention
  - Mutilated DAGs as Causal Networks
  - Casual Probability Calculations and "Causal Effects" (Pearl's Theory)
- Two Key (and Controversial) Assumptions of Bayesian Network Modeling
  - The Causal Markov Condition (PCC and Faithfulness)
  - The Stability Assumption (invariance of independencies)

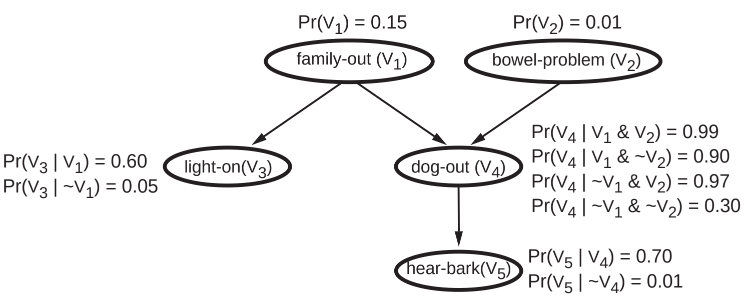
Conditional Independence Structures and DAGs (Again)

- There are 3 types of connections between triples of variables X, Y, and Z.



- In the linear case, we have $(X \perp\!\!\!\perp Y \mid Z)$. In the converging case, we have $(X \not\perp\!\!\!\perp Y \mid Z)$, and in the diverging case, we have $(X \perp\!\!\!\perp Y \mid Z)$.
- A path p is *d-separated* (or *blocked*) by (a *set*) $\mathbf{Z}$ iff either:
  1. p contains an $i \rightarrow m \rightarrow j$ or an $i \leftarrow m \rightarrow j$ such that $m \in \mathbf{Z}$.
  2. p contains an $i \rightarrow m \leftarrow j$ such that $m \notin \mathbf{Z}$ and no descendant of m is in $\mathbf{Z}$.
- Iff $\mathbf{Z}$ blocks every path from X to Y, then $\mathbf{Z}$ *d-separates X from Y* and $(X \perp\!\!\!\perp Y \mid \mathbf{Z})$.

Evidential Probability Calculations in Networks



- In Charniak's example, we may compute all of the unspecified probabilities recursively, beginning from V_3, which only depends on V_1, as follows.
- From the Markovian assumptions implicit in the network, we know that V_3 depends only on V_1. Thus, we can compute the probability of V_3 as follows: $\Pr(V_3) = \Pr(V_3 \mid V_1) \cdot \Pr(V_1) + \Pr(V_3 \mid \neg V_1) \cdot \Pr(\neg V_1) = .6 \cdot .15 + .05 \cdot .85 = .1325$.

- $\Pr(V_4)$ is tougher. For this, we can use the same total probability expansion: $\Pr(V_4) = \Pr(V_4 \mid V_1) \cdot \Pr(V_1) + \Pr(V_4 \mid \neg V_1) \cdot \Pr(\neg V_1)$. But, we must also expand $\Pr(V_4 \mid V_1) = \Pr(V_4 \mid V_1 \,\&\, V_2) \cdot \Pr(V_2 \mid V_1) + \Pr(V_4 \mid V_1 \,\&\, \neg V_2) \cdot \Pr(\neg V_2 \mid V_1)$. From the Markovian assumptions implicit in the network, we know that $\Pr(V_2 \mid V_1) = \Pr(V_2) = .01$. So, $\Pr(V_4 \mid V_1) = .99 \cdot .01 + .90 \cdot .99 = .9009$. A similar computation yields $\Pr(V_4 \mid \neg V_1) = .97 \cdot .01 + .3 \cdot .99 = .3067$. Finally, we have $\Pr(V_4) = .9009 \cdot .15 + .3067 \cdot .85 = .39583$. I'd rather ask Audrey's applet!
- $\Pr(V_5)$ is now easy, since it only depends on V_4. So, we have: $\Pr(V_5) = \Pr(V_5 \mid V_4) \cdot \Pr(V_4) + \Pr(V_5 \mid \neg V_4) \cdot \Pr(\neg V_4) = .7 \cdot .39583 + .01 \cdot .60417 = .283123$.
- Now that we have all the unconditional probabilities of the V_i, computing all the remaining probabilities will involve applications of Bayes' theorem (together with Markovian knowledge). Consider $\Pr(V_1 \mid V_3 \,\&\, V_5)$ (which Charniak says is 0.5, and which we computed using Audrey's Java Applet last time).
- This computation is non-trivial, but it's just a matter of exploiting conditional independencies and applying Bayes' theorem. This is a good exercise in probabilistic reasoning! It's a good thing we have programs to do these for us!

Traditional PC and Evidential Bayesian Networks

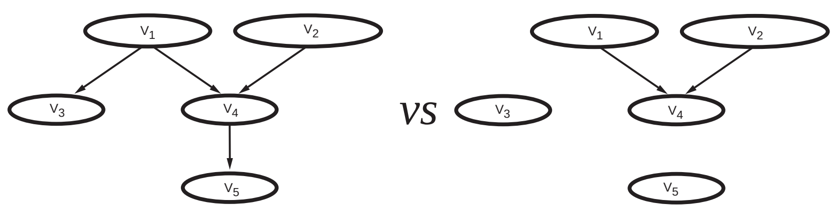
- Given a (dichotomous, evidential) Bayesian network structure, it might *seem* straightforward to determine if X causes Y in the traditional PC sense.
- One might *think* that one (*qua* traditional PCer) could say that X is causally relevant to Y (in a graph model $\mathcal{G}$) just in case $(X \not\perp Y \mid \mathbf{K})$, where $\mathbf{K}$ is the set of variables (in $\mathcal{G}$) such that each member V of $\mathbf{K}$ is such that $V \to Y$ and $X \nrightarrow V$.
- Is this OK? I can't think of any obvious counterexamples, can you?
- How about Hitchcock's refinement of the traditional PC theory?
- On his account, we always are supposed to ask if $X = x$, as opposed to $X = x'$, is a cause of Y. Specifically, we are supposed to compare $\Pr(Y \mid X = x)$ and $\Pr(Y \mid X = x')$. But, presumably, these Pr's are to be conditionalized on $\mathbf{K}$.
- One might then say that the "causal effect" (in the traditional PC sense) of X on Y (in $\mathcal{G}$) is captured by the conditional distribution function $\Pr(Y \mid X = x \,\&\, \mathbf{K})$.
- Pearl does *not* make this kind of move (neither does Hitchcock *now*).

Causal Bayesian Networks I

- Pearl (in his recent 2000 book) makes a significant addition to his (1988) evidential Bayesian network theory (which some have used for "causal talk").
- What's new is the notion of an *intervention* (along the lines described in Woodward's SEP piece). Pearl distinguishes the evidential probability $\Pr(Y = y \mid X = x)$ from the causal probability $\Pr(Y = y \mid do(X = x))$.
- The latter [also written $\Pr(y \mid \hat{x})$] is intended to be the distribution on Y induced by and *intervention* which *holds fixed* (come what may) X at the value x.
- As it turns out, there is a very elegant characterization of how to compute such probabilities within Pearl's causal Bayesian networks (as DAGs).
- The techniques will be largely the same as before (Audrey's applet to the rescue again!), but we will now apply them to *different graphs*.
- Calculating $\Pr(y \mid x)$ is just the kind of thing we saw with Charniak's network. Calculating $\Pr(y \mid \hat{x})$ requires us to *mutilate* (Pearl's term) the original graph . . .

Causal Bayesian Networks II

- We have $\Pr(v_1 \mid v_3 \,\&\, v_5) = 0.5 > \Pr(v_1) = 0.15$. This is just the probability that the family is out, *on the supposition that* the light is on and the dog is barking.
- On the other hand, $\Pr(v_1 \mid \hat{v}_3 \,\&\, \hat{v}_5) \approx 0.08 < \Pr(v_1) = 0.15$. So, while the *evidential* impact of $v_3 \,\&\, v_5$ on v_1 is positive, its *causal* impact is nil.



- Mutilated graphs (right) are obtained by deleting from evidential graphs (left) all arrows coming into variables whose values are being set by interventions.
- The effect of "mutilation" is to simply *delete* the probabilities [$\Pr(x_i \mid pa_i)$], corresponding to the deleted arrow(s) [$PA_i \to X_i$], from the factorization(s).

Causal Bayesian Networks III

- One of Pearl's key assumptions is that the value of each variable V_j in a DAG $\mathcal{G}$ is determined by the values pa_j of its parents PA_j in $\mathcal{G}$, *modulo* an "error term" u which is there only because of latent or hidden variables that we have yet to discover, and are not included in the model because of our ignorance.
- More formally, this assumption postulates a system of *structural equations*: $v_j = f_j(pa_j, u_j)$, for $j = 1, \ldots, n$. Note that Sewall Wright's *path analysis* is just a *special case* of this set-up, in which the f_j are in a certain *linear* form.
- The "error terms" u_j are assumed to be *mutually independent*, to ensure that the "error distribution" $\Pr(v)$ induced by the u_j is Markovian. If all relevant factors have been included in the model, then $u_j = 0$, and $\Pr(v) \in \{0, 1\}$. Probability is merely a measure of ignorance of deterministic functional (causal) structure.
- DAGs, joint (Markovian) distributions, and structural equations of this kind are just three different ways of characterizing the same underlying structure. But, the structural equations are most informative and straightforward representation.

Causal Bayesian Networks IV

- Pearl gives the following definition of the "causal effect" of set X on set Y, in terms of the underlying structural equation model.

  Given two disjoint sets of variables, X and Y (in a model/graph $\mathcal{G}$), the causal effect of X on Y (in $\mathcal{G}$), denoted as $\Pr(y \mid \hat{x})$ or $\Pr(y \mid do(x))$, is a function from X to the space of probability distributions on Y. For each realization x of X, $\Pr(y \mid \hat{x})$ gives the probability of $Y = y$ induced by deleting from the structural equation model all equations corresponding to variables in X and substituting $X = x$ in the remaining equations.

- This is equivalent to taking as our new DAG model (and corresponding $\Pr(v)$) $\mathcal{G}'$ the subgraph of $\mathcal{G}$ in which all arrows into variables in X have been pruned.
- Pearl's "causal effect" is similar to Hitchcock's, but more general in two ways.
  - Pearl's "causal effect" relation is a a relation between *sets of* variables, not just between variables.
  - Pearl's "causal effect" relation is a relation between *variables*, not just between variables and *propositions*.
- What kind of thing is a *set of variables*? Yet another "level" of causation?

Causal Bayesian Networks V

- A 2nd assumption is the *Causal Markov Condition* (CM), which has 2 parts.

CM1. If X and Y are probabilistically dependent (in $\mathcal{G}$), then either X causes Y or Y causes X or there exists a common cause Z (in $\mathcal{G}$) of X and Y.

CM2. If set of parents of X [$PA(x)$, in $\mathcal{G}$] is non-empty, then $(X \perp\!\!\!\perp V \mid PA(x))$, for all V except the effects of X (in $\mathcal{G}$).

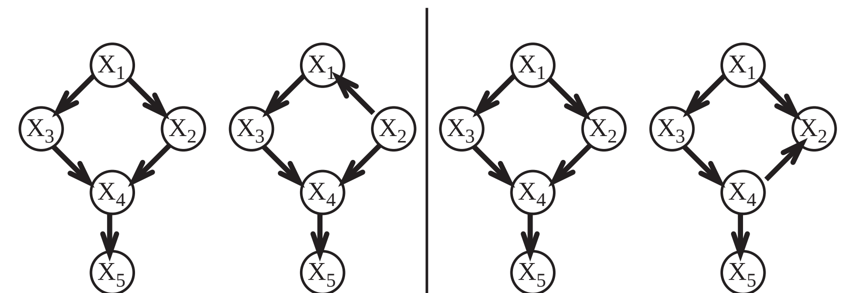
- CM1 is an instance what has become known in the philosophical literature as "the principle of the common cause". This was assumed by Reichenbach (1956), and by many people since. The converse (also assumed by Reichenbach and many others) is called the "faithfulness" condition.
- CM2 is also something that has been assumed by many (including I.J. Good 1960). I would say that both assumptions are controversial. But, CM1 has received more critical attention in philosophy. Pat Suppes is one of the vocal critics of CM2. Hopefully, he'll be talking to us about that in a couple of weeks.
- In the next couple of weeks, we'll be reading *a lot* about these!

Causal Bayesian Networks VI

- A third key assumption is what Pearl calls "Stability". Each causal model consists of a set of structural equations, which come along with a set of parameters Θ that fix the functions f_i and the error distributions u_i.
- Each causal model determines a set of conditional independence constraints, represented implicitly in the DAG. The *stability* condition requires that all of the probabilistic independence relations implied by the model should be *invariant* (stable) across (small) perturbations to the parameters of the model.
- Another way of saying this is that the parameters of the models should not *themselves* be functionally (equationally) related to each other.
- One "argument" for this is that the probability that such functional relations will hold between parameters is zero, if the errors are mutually independent.
- Freedman (see readings) cleverly points out that the independence relations *themselves* can be thought of as functional relations between parameters in the model. But, presumably, *this* is no reason to think *they* are unstable.

Causal Bayesian Networks VII

- Verma & Pearl proved that two DAGs are (passively) *observationally equivalent* if (i) they have the same skeletons, and (ii) the same sets of v-structures, i.e., two converging arrows whose tails are not connected by an arrow.
- Using this criterion, we can see that (a) and (b) are *indistinguishable* (from *passive* probabilistic information alone), but (c) and (d) are not.



(a) (b) (c) (d)