In what follows [Dav] refers to Davison's book, and [Jen] refers to Jensen's 1996 book on Bayesian networks.

1. Show that for a DAG, the local directed Markov property (DL) implies the ordered directed Markov property (DO).

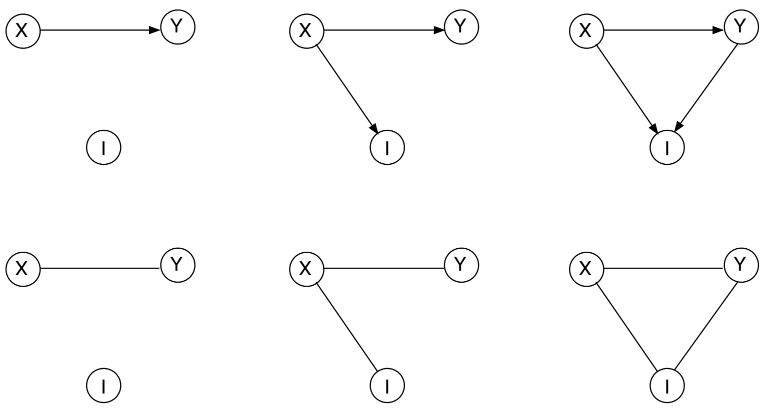
Solution: By definition, a distribution $P(X)$ is said to obey (DL) if $X_v \perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)}$ for all $v \in V$, where $\mathrm{nd}(v)$ is the set of non-descendants of v, and to obey (DO) if $X_v \perp X_{\mathrm{pr}(v)} \mid X_{\mathrm{pa}(v)}$ for all $v \in V$, where $\mathrm{pr}(v)$ is the set of predecessors of v with respect to some given well-ordering. By the definition of a well-ordering, all the descendants of v have larger numbers than v, while all nodes in $\mathrm{pr}(v)$ have smaller numbers than v. So $\mathrm{pr}(v) \subseteq \mathrm{nd}(v)$ for all $v \in V$, which shows that the variables $X_{\mathrm{pr}(v)}$ are included in $X_{\mathrm{nd}(v)}$ for all $v \in V$. Therefore, (DL) implies (DO). $\square$

2. Data-Missing Mechanisms. (This is Exercise 6.2.10 in [Dav].) Suppose our goal is inference for a parameter θ based on data that would ideally consist of n independent pairs (X, Y), but that some values of Y are missing, as shown by an indicator variable I. Thus the data on an individual have form $(x, y, 1)$ or $(x, ?, 0)$. We suppose that although the missing mechanism $\mathrm{Prob}(I = 0 \mid x, y)$ may depend on x and y, it does not involve θ. There are now three possibilities:

   - data are *missing completely at random*, that is $\mathrm{Prob}(I = 0 \mid x, y) = \mathrm{Prob}(I = 0)$ is independent of both x and y;
   - data are *missing at random*, that is $\mathrm{Prob}(I = 0 \mid x, y) = \mathrm{Prob}(I = 0 \mid x)$ depends on x but not on y;
   - there is *non-ignorable non-response*, meaning that $\mathrm{Prob}(I = 0 \mid x, y)$ depends on y and possibly also on x.

   Write down the DAG and its moral graph for X, Y and I under the missing data models described above. Use them to give an equation-free explanation of the differences among the models and of their consequences.

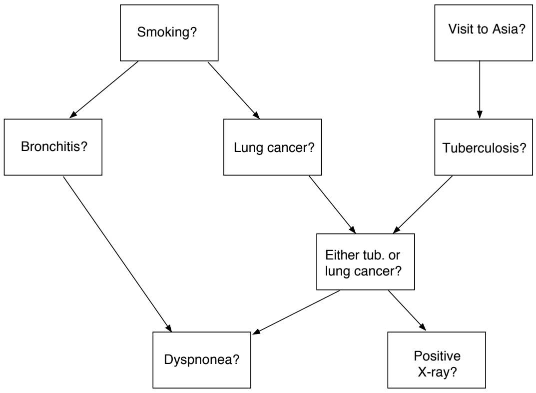
Solution: The DAGs for the three models are shown below (in the first row) with their moral graphs (in the second row), which coincide with the undirected versions of the DAGs.

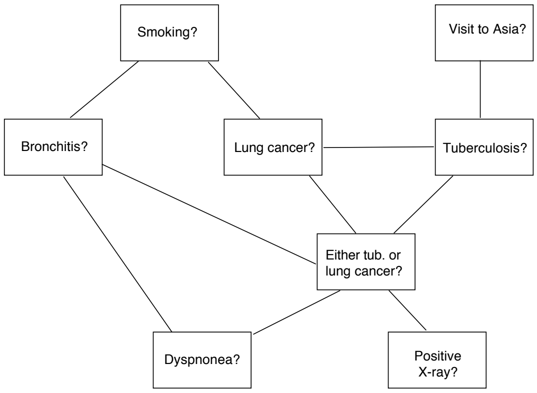


My explanation of the differences of the models and of their consequences are as follows. Under the first two models, given X, Y is conditionally independent of I, so I carries no information about Y. Since our goal is to infer the distribution of (X, Y) and not that of I given X, we can base inference on the observed data of the form (x, y) or (x) if y is missing, ignoring the value of I. Under the third model, given X, Y is not conditionally independent of I, so I carries information about Y and is non-ignorable, and we need to know the exact missing mechanism $\mathrm{Prob}(I = 0 \mid x, y)$ in order to infer the probabilistic relation between X and Y. $\square$

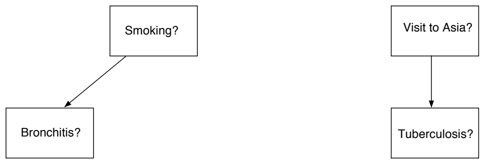
3. Consider the Asia network in the book by Cowell et al. (Chap. 2, p. 20). Write down the moral graph. Is the variable 'Bronchitis' independent of 'Tuberculosis?' without observing anything? Are they independent of each other after receiving the X-ray result? And how about after knowing both the X-ray result and smoking history?

Solution: The Asia network (left) and its moral graph (right):

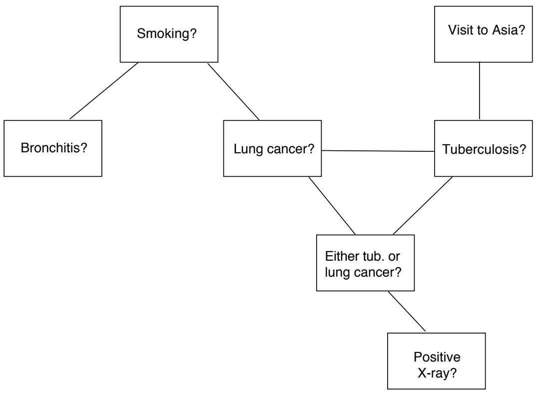




Without observing anything, the variable 'Bronchitis' is independent of 'Tuberculosis'. This can be seen from the subgraph induced by the minimal ancestral set containing the two nodes 'Bronchitis' and 'Tuberculosis,' as shown on the right. The moral graph coincides with the undirected version of the subgraph. The two nodes 'Bronchitis' and 'Tuberculosis' are disconnected, therefore by the global directed Markov property (DG), the two variables are independent.

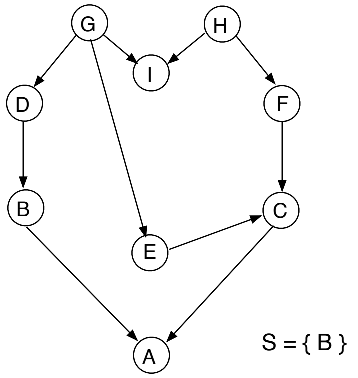


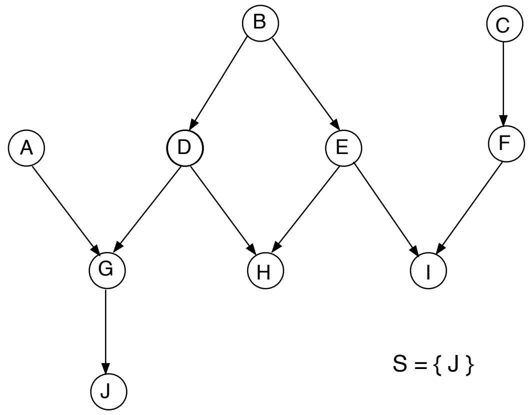
The variables 'Bronchitis' and 'Tuberculosis' are conditionally dependent after receiving the X-ray result. This can be seen from the moral graph of the subgraph induced by the minimal ancestral set containing the three nodes: 'Bronchitis,' 'Tuberculosis' and 'Positive X-ray,' as shown below. In this moral graph, 'Bronchitis' and 'Tuberculosis' are not separated by 'Positive X-ray,' therefore we cannot assert using (DG) that the two variables are conditionally independent given 'Positive X-ray.'



But 'Bronchitis' and 'Tuberculosis' are conditionally independent after receiving also the smoking history. Again, this can be seen from the moral graph of the subgraph induced by the minimal ancestral set containing the four nodes: 'Bronchitis,' 'Tuberculosis,' 'Positive X-ray,' and 'Smoking,' which is the same graph as shown above. In this moral graph, 'Bronchitis' and 'Tuberculosis' are separated by 'Smoking.' Therefore by (DG), the two variables are conditionally independent given X-ray result and smoking history. □

4. (This is Exercise 2.2 in [Jen].) For the two graphs below determine which variables are d-connected to (i.e., not d-separated from) A given S.

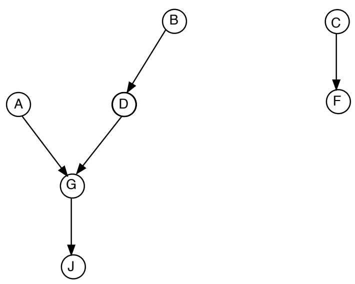


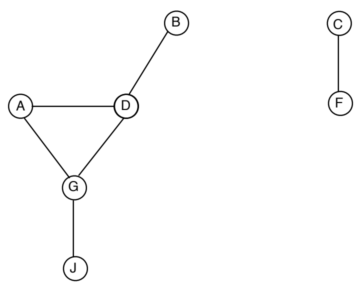


Solution: For the first graph, all the variables, $C, D, \ldots, I$, are d-connected to A given $S = B$, because there is an active trail from each of these nodes to A. For example, the trail (D, G, E, C, A) circumvents B and does not have converging arrows at any of its nodes, therefore it is active. And so is the trail (I, H, F, C, A). From these two active trails, we can find pieces of trails, which are themselves active trails, linking each of $C, D, \ldots, I$ to A.

For the second graph, given $S = J$, F and C are d-separated from A, while the rest of the variables are d-connected to A. Since there are very few trails, this can be argued directly using the definition of d-separation and active trails as before: (H, D, G, A) and (I, E, B, D, G, A) are active trails, so all the nodes on them are d-connected to A. Any trail from F or C to A must go through (F, I, E) with a converging connection at I; but I is not in S and has no descendants, so the trail is blocked. This shows F and C are d-separated from A.

Alternatively, we can also use (DG) to show the d-separation of F and C from A. From the subgraph $G_{\mathrm{An}(\{A,J,F,C\})}$ induced by the minimal ancestral set $\mathrm{An}(\{A, J, F, C\})$ containing $\{A, J, F, C\}$, and its moral graph $\left(G_{\mathrm{An}(\{A,J,F,C\})}\right)^m$:





we see that $\{F, C\}$ are disconnected from $\{A, J\}$. Therefore, by (DG) and its equivalence to d-separation, F and C are d-separated from A given J. $\square$