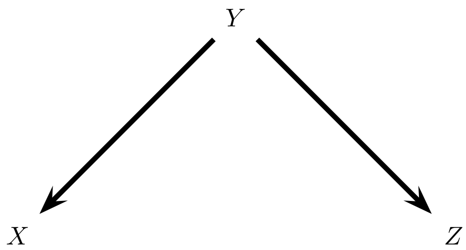

Figure 90: A directed graph with vertices $V = \{X, Y, Z\}$ and edges $E = \{(Y, X), (Y, Z)\}$.

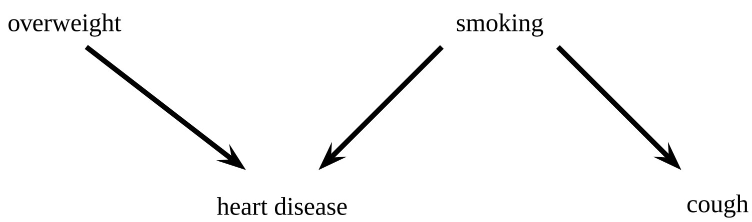

Figure 91: DAG for Example 48.1.

48 Directed Graphs

A directed graph consists of a set of nodes with arrows between some nodes. An example is shown in Figure 90.

Formally, a **directed graph** $\mathcal{G}$ consists of a set of vertices V and an edge set E of ordered pairs of vertices. For our purposes, each vertex will correspond to a random variable. If $(X, Y) \in E$ then there is an arrow pointing from X to Y. See Figure 90.

If an arrow connects two variables X and Y (in either direction) we say that X and Y are **adjacent.** If there is an arrow from X to Y then X is a **parent** of Y and Y is a **child** of X. The set of all parents of X is denoted by π_X or $\pi(X)$. A **directed path** between two variables is a set of arrows all pointing in the same direction linking one variable to the

other such as:

A sequence of adjacent vertices staring with X and ending with Y but ignoring the direction of the arrows is called an **undirected path.** The sequence $\{X, Y, Z\}$ in Figure 90 is an undirected path. X is an **ancestor** of Y if there is a directed path from X to Y (or $X = Y$). We also say that Y is a **descendant** of X.

A configuration of the form:

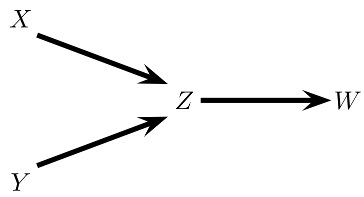

Figure 92: Another DAG.

is called a **collider** at Y. A configuration not of that form is called a **non-collider**, for example,

or

The collider property is path dependent. In Figure 96, Y is a collider on the path $\{X, Y, Z\}$ but it is a non-collider on the path $\{X, Y, W\}$. When the variables pointing into the collider are not adjacent, we say that the collider is **unshielded**. A directed path that starts and ends at the same variable is called a **cycle.** A directed graph is **acyclic** if it has no cycles. In this case we say that the graph is a **directed acyclic graph** or **DAG**. From now on, we only deal with acyclic graphs.

Let $\mathcal{G}$ be a DAG with vertices $V = (X_1, \ldots, X_k)$.

If $\mathbb{P}$ is a distribution for V with probability function f, we say that $\mathbb{P}$ is **Markov to** $\mathcal{G}$, or that $\mathcal{G}$ **represents** $\mathbb{P}$, if

$$f(v) = \prod_{i=1}^{k} f(X_i \mid \pi_i) \tag{329}$$

where π_i are the parents of X_i. The set of distributions represented by $\mathcal{G}$ is denoted by $M(\mathcal{G})$.

48.1 Example. Figure 91 shows a DAG with four variables. The probability function for this example factors as

$$\begin{aligned} & f(\text{overweight}, \text{smoking}, \text{heart disease}, \text{cough}) \\ & \quad = \; f(\text{overweight}) \times f(\text{smoking}) \\ & \quad \times \; f(\text{heart disease} \mid \text{overweight}, \text{smoking}) \\ & \quad \times \; f(\text{cough} \mid \text{smoking}). \quad \blacksquare \end{aligned}$$

48.2 Example. For the DAG in Figure 92, $\mathbb{P} \in M(\mathcal{G})$ if and only if its probability function f has the form

$$f(x, y, z, w) = f(x) f(y) f(z \mid x, y) f(w \mid z). \quad \blacksquare$$

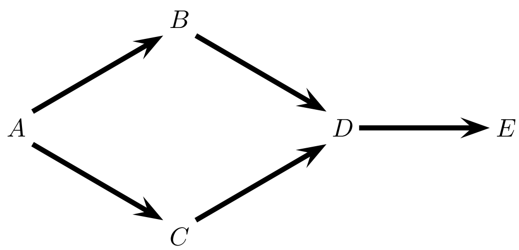

Figure 93: Yet another DAG.

The following theorem says that $\mathbb{P} \in M(\mathcal{G})$ if and only if the **Markov Condition** holds. Roughly speaking, the Markov Condition means that every variable W is independent of the "past" given its parents.

48.3 Theorem. *A distribution $\mathbb{P} \in M(\mathcal{G})$ if and only if the following* **Markov Condition** *holds: for every variable W,*

$$W \amalg \widetilde{W} \mid \pi_W \tag{330}$$

where $\widetilde{W}$ denotes all the other variables except the parents and descendants of W.

48.4 Example. In Figure 92, the Markov Condition implies that

$$X \amalg Y \quad \text{and} \quad W \amalg \{X, Y\} \mid Z. \quad \blacksquare$$

48.5 Example. Consider the DAG in Figure 93. In this case probability function must factor like

$$f(a, b, c, d, e) = f(a)f(b|a)f(c|a)f(d|b, c)f(e|d).$$

The Markov Condition implies the following independence relations:

$$D \amalg A \mid \{B, C\}, \quad E \amalg \{A, B, C\} \mid D \quad \text{and} \quad B \amalg C \mid A \quad \blacksquare$$

The Markov Condition allows us to list some independence relations implied by a DAG. These relations might imply other independence relations. Consider the DAG in Figure 94. The Markov Condition implies:

$$X_1 \amalg X_2, \qquad X_2 \amalg \{X_1, X_4\}, \qquad X_3 \amalg X_4 \mid \{X_1, X_2\},$$

$$X_4 \amalg \{X_2, X_3\} \mid X_1, \qquad X_5 \amalg \{X_1, X_2\} \mid \{X_3, X_4\}$$

It turns out (but it is not obvious) that these conditions imply that

$$\{X_4, X_5\} \amalg X_2 \mid \{X_1, X_3\}.$$

How do we find these extra independence relations? The answer is "d-separation" which means "directed separation." d-separation can be summarized by three rules. Consider the four DAG's in Figure 95 and the DAG in Figure 96. The first 3 DAG's in Figure 95 have no colliders. The DAG in the lower right of Figure 95 has a collider. The DAG in Figure 96 has a collider with a descendant.

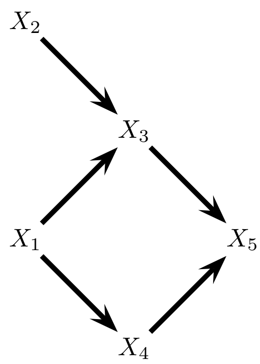

Figure 94: And yet another DAG.

Figure 95: The first three DAG's have no colliders. The fourth DAG in the lower right corner has a collider at Y.

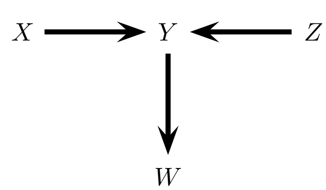

Figure 96: A collider with a descendant.

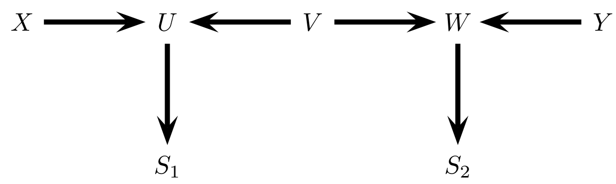

Figure 97: d-separation explained.

The Rules of d-Separation

Consider the DAGs in Figures 95 and 96.

1. When Y is not a collider, X and Z are **d-connected**, but they are **d-separated** given Y.
2. If X and Z collide at Y, then X and Z are **d-separated**, but they are **d-connected** given Y.
3. Conditioning on the descendant of a collider has the same effect as conditioning on the collider. Thus in Figure 96, X and Z are **d-separated** but they are **d-connected** given W.

Here is a more formal definition of d-separation. Let X and Y be distinct vertices and let W be a set of vertices not containing X or Y. Then X and Y are **d-separated given** W if there exists no undirected path U between X and Y such that (i) every collider on U has a descendant in W, and (ii) no other vertex on U is in W. If A, B, and W are distinct sets of vertices and A and B are not empty, then A and B are d-separated given W if for every $X \in A$ and $Y \in B$, X and Y are d-separated given W. Sets of vertices that are not d-separated are said to be d-connected.

48.6 Example. Consider the DAG in Figure 97. From the d-separation rules we conclude that:
X and Y are d-separated (given the empty set);
X and Y are d-connected given $\{S_1, S_2\}$;
X and Y are d-separated given $\{S_1, S_2, V\}$.

48.7 Theorem. [8] *Let A, B, and C be disjoint sets of vertices. Then $A \amalg B \mid C$ if and only if A and B are d-separated by C.*

48.8 Example. The fact that conditioning on a collider creates dependence might not seem intuitive. Here is a whimsical example from Jordan (2003). that makes this idea more palatable. Your friend appears to be late for a meeting with you. There are two explanations: she was abducted by aliens or you forgot to set your watch ahead one hour for daylight savings time. (See Figure 98.) Aliens and Watch are blocked by a collider which implies they are marginally independent. This seems reasonable since — before we know anything about your friend being late — we would expect these variables to be independent. We would also expect that $\mathbb{P}(\mathsf{Aliens} = \mathsf{yes}|\mathsf{Late} = \mathsf{yes}) > \mathbb{P}(\mathsf{Aliens} = \mathsf{yes})$; learning that your friend is late certainly increases the probability that she was abducted. But when we learn that you forgot to set your watch properly, we would lower the chance that your friend was abducted. Hence, $\mathbb{P}(\mathsf{Aliens} = \mathsf{yes}|\mathsf{Late} = \mathsf{yes}) \neq \mathbb{P}(\mathsf{Aliens} = \mathsf{yes}|\mathsf{Late} = \mathsf{yes}, \mathsf{Watch} = \mathsf{no})$. Thus, Aliens and Watch are dependent given Late. ■

48.9 Example. Consider the DAG in Figure 91. In this example, overweight and smoking are marginally independent but they are dependent given heart disease. ■

Graphs that look different may actually imply the same independence relations. If $\mathcal{G}$ is a DAG, we let $\mathcal{I}(\mathcal{G})$ denote all the independence statements implied by $\mathcal{G}$. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ for the same variables V are **Markov equivalent** if $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$. Given a DAG $\mathcal{G}$, let $\mathrm{skeleton}(\mathcal{G})$ denote the undirected graph obtained by replacing the arrows with undirected edges.

[8] We implicitly assume that $\mathbb{P}$ is **faithful** to $\mathcal{G}$ which means that $\mathbb{P}$ has no extra independence relations other than those logically implied by the Markov Condition.

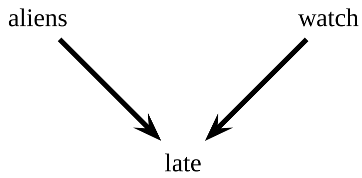

Figure 98: Jordan's alien example (Example 48.8). Was your friend kidnapped by aliens or did you forget to set your watch?

48.10 Theorem. *Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if (i)* $\text{skeleton}(\mathcal{G}_1) = \text{skeleton}(\mathcal{G}_2)$ *and (ii) $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders.*

48.11 Example. The first three DAGs in Figure 95 are Markov equivalent. The DAG in the lower right of the Figure is not Markov equivalent to the others. ■

49 Estimation for DAGs

Estimating a DAG structure is harder than an undirected graph. For the sparse graphs, the PC algorithm due to Spirtes, Glymour and Scheines is the fastest algorithm I know of. Here, we will consider the simpler case where there is a known ordering $\preceq$ on the variables. An example is time order. Without loss of generality, assume that $V = \{1, \ldots, d\}$ has been ordered according to $\preceq$. Thus, $i \preceq j$ if and only if $i \leq j$.

For continuous, Gaussian variables, and small graphs we can use SIN. For every $i < j$ we test

$$H_0 : \rho^{ij}_{\leftarrow} = 0 \quad \text{versus} \quad H_1 : \rho^{ij}_{\leftarrow} \neq 0. \tag{331}$$

Here, $\rho^{ij}_{\leftarrow}$ denotes the partial correlation of X_i and X_j given $\{1, \ldots, j\} - \{i, j\}$.

49.1 Example. This example, from Spirtes et al (2000) involves data on publishing productivity. The variables are: sex, ability, GPQ (graduate program quality), preprod (preliminary measure of productivity), QFJ (quality of first job), pubs (publication rates), and cites (citation rates). These are essentially time ordered.

```
> library(SIN)
> postscript("dagsin.ps",horizontal=FALSE)
> data(pubprod)
> attach(pubprod)
> m = pubprod$cor
> print(dimnames(m)[[1]])
[1] "ability" "GPQ"     "preprod" "QFJ"     "sex"     "cites"   "pubs"
> n = pubprod$n
> o = c(5,1,2,3,4,7,6)
> ### sex < ability < GPQ < pre < QFJ < pubs < cites
> m = m[o,o]
> p = sinDAG(1:7,m,n)
> plotDAGpvalues(p)
```

```
> G = getgraph(p,.05,type="DAG")
> print(G)
        sex ability GPQ preprod QFJ pubs cites
sex       0       0   0       0   0    1     0
ability   0       0   1       1   0    0     0
GPQ       0       0   0       0   1    0     0
preprod   0       0   0       0   0    0     1
QFJ       0       0   0       0   0    1     0
pubs      0       0   0       0   0    0     1
cites     0       0   0       0   0    0     0
```

The p-values are plotted in Figure 99. ■

An alternative is simply to do the following:
regress X_2 on X_1
regress X_3 on X_1, X_2
regress X_4 on X_1, X_2, X_3
etc.
One could test for significant effects or, when d is large, use the lasso.

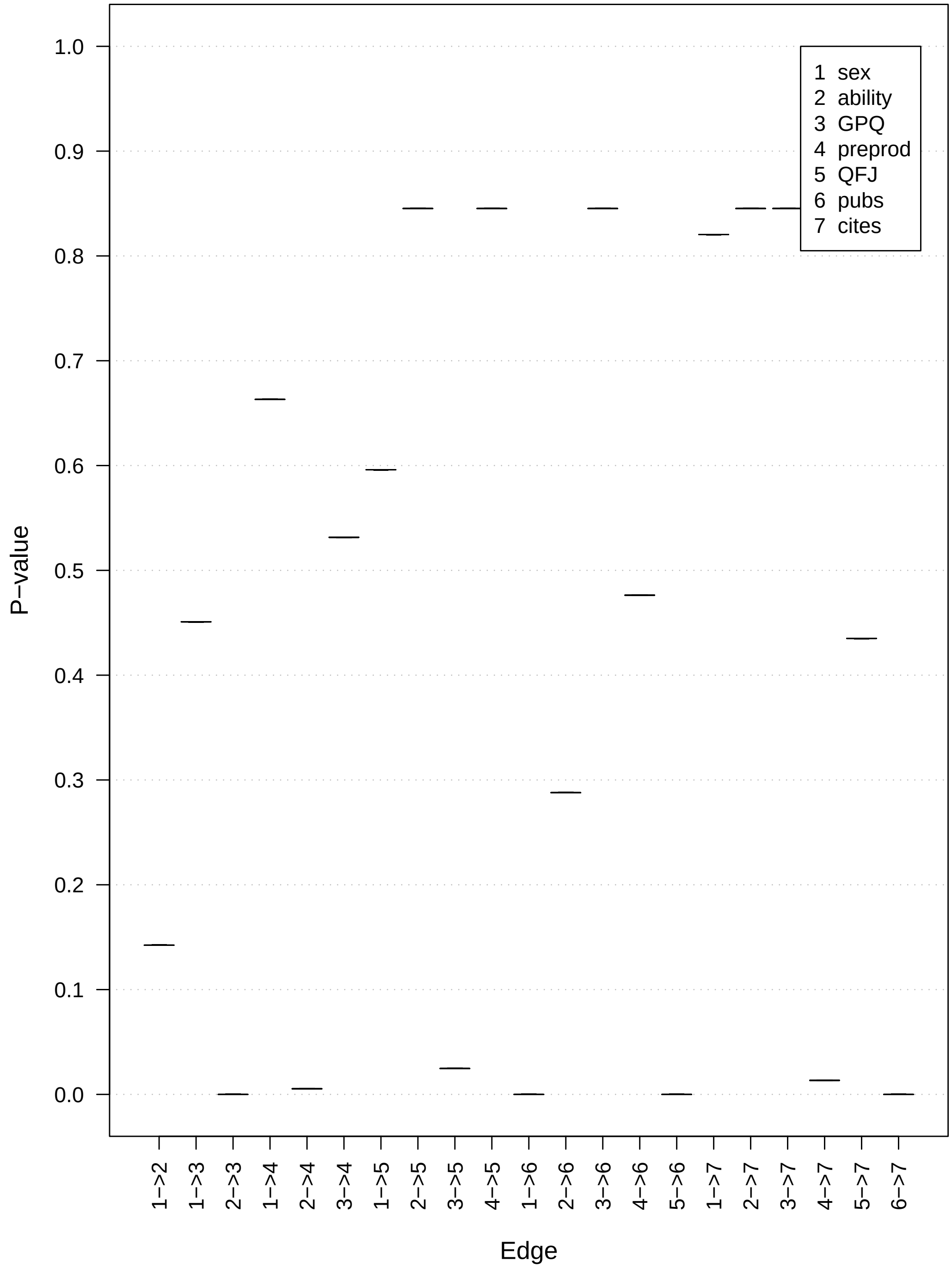

Figure 99: p-values for DAG example