18

Undirected Graphs

Undirected graphs are an alternative to directed graphs for representing independence relations. Since both directed and undirected graphs are used in practice, it is a good idea to be facile with both. The main difference between the two is that the rules for reading independence relations from the graph are different.

18.1 Undirected Graphs

An **undirected graph** $\mathcal{G} = (V, E)$ has a finite set V of **vertices (or nodes)** and a set E of **edges (or arcs)** consisting of pairs of vertices. The vertices correspond to random variables $X, Y, Z, \ldots$ and edges are written as unordered pairs. For example, $(X, Y) \in E$ means that X and Y are joined by an edge. An example of a graph is in Figure 18.1.

Two vertices are **adjacent**, written $X \sim Y$, if there is an edge between them. In Figure 18.1, X and Y are adjacent but X and Z are not adjacent. A sequence $X_0, \ldots, X_n$ is called a **path** if $X_{i-1} \sim X_i$ for each i. In Figure 18.1, X, Y, Z is a path. A graph is **complete** if there is an edge between every pair of vertices. A subset $U \subset V$ of vertices together with their edges is called a **subgraph.**

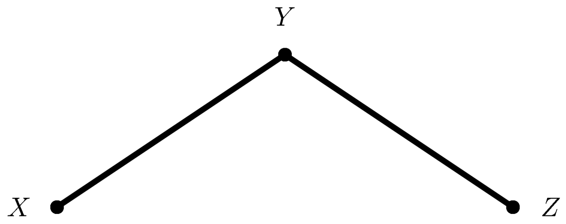

FIGURE 18.1. A graph with vertices $V = \{X, Y, Z\}$. The edge set is $E = \{(X, Y), (Y, Z)\}$.

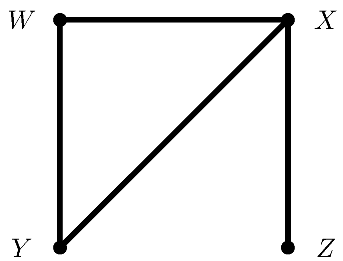

FIGURE 18.2. $\{Y, W\}$ and $\{Z\}$ are separated by $\{X\}$. Also, W and Z are separated by $\{X, Y\}$.

If A, B and C are three distinct subsets of V, we say that C **separates** A **and** B if every path from a variable in A to a variable in B intersects a variable in C. In Figure 18.2 $\{Y, W\}$ and $\{Z\}$ are separated by $\{X\}$. Also, W and Z are separated by $\{X, Y\}$.

18.2 Probability and Graphs

Let V be a set of random variables with distribution $\mathbb{P}$. Construct a graph with one vertex for each random variable in V. Omit the edge between a pair of variables if they are independent given the rest of the variables:

$$\text{no edge between } X \text{ and } Y \iff X \amalg Y | \text{rest}$$

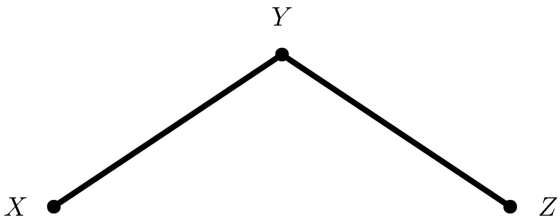

FIGURE 18.3. $X \amalg Z|Y$.

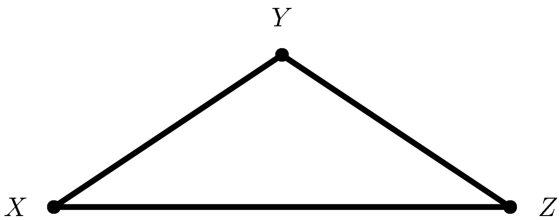

FIGURE 18.4. No implied independence relations.

where "rest" refers to all the other variables besides X and Y. The resulting graph is called a **pairwise Markov graph**. Some examples are shown in Figures 18.3, 18.4, 18.5, and 18.6.

The graph encodes a set of pairwise conditional independence relations. These relations imply other conditional independence relations. How can we figure out what they are? Fortunately, we can read these other conditional independence relations directly from the graph as well, as is explained in the next theorem.

18.1 Theorem. *Let $\mathcal{G} = (V, E)$ be a pairwise Markov graph for a distribution $\mathbb{P}$. Let A, B and C be distinct subsets of V such that C separates A and B. Then $A \amalg B|C$.*

18.2 Remark. If A and B are not connected (i.e., there is no path from A to B) then we may regard A and B as being separated by the empty set. Then Theorem 18.1 implies that $A \amalg B$.

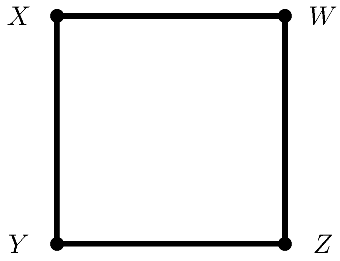

FIGURE 18.5. $X \amalg Z|\{Y, W\}$ and $Y \amalg W|\{X, Z\}$.

FIGURE 18.6. Pairwise independence implies that $X \amalg Z|\{Y, W\}$. But is $X \amalg Z|Y$?

The independence condition in Theorem 18.1 is called the **global Markov property.** We thus see that the pairwise and global Markov properties are equivalent. Let us state this more precisely. Given a graph $\mathcal{G}$, let $M_{\text{pair}}(\mathcal{G})$ be the set of distributions which satisfy the pairwise Markov property: thus $\mathbb{P} \in M_{\text{pair}}(\mathcal{G})$ if, under $\mathbb{P}$, $X \amalg Y|\text{rest}$ if and only if there is no edge between X and Y. Let $M_{\text{global}}(\mathcal{G})$ be the set of distributions which satisfy the global Markov property: thus $\mathbb{P} \in M_{\text{pair}}(\mathcal{G})$ if, under $\mathbb{P}$, $A \amalg B|C$ if and only if C separates A and B.

18.3 Theorem. *Let $\mathcal{G}$ be a graph. Then, $M_{\text{pair}}(\mathcal{G}) = M_{\text{global}}(\mathcal{G})$.*

Theorem 18.3 allows us to construct graphs using the simpler pairwise property and then we can deduce other independence relations using the global Markov property. Think how hard this would be to do algebraically. Returning to 18.6, we now see that $X \amalg Z|Y$ and $Y \amalg W|Z$.

18.4 Example. Figure 18.7 implies that $X \amalg Y$, $X \amalg Z$ and $X \amalg (Y, Z)$. ■

18.5 Example. Figure 18.8 implies that $X \amalg W|(Y, Z)$ and $X \amalg Z|Y$. ■

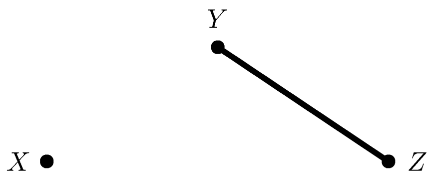

FIGURE 18.7. $X \amalg Y$, $X \amalg Z$ and $X \amalg (Y, Z)$.

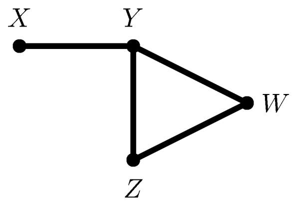

FIGURE 18.8. $X \amalg W|(Y, Z)$ and $X \amalg Z|Y$.

18.3 Cliques and Potentials

A **clique** is a set of variables in a graph that are all adjacent to each other. A set of variables is a **maximal clique** if it is a clique and if it is not possible to include another variable and still be a clique. A **potential** is any positive function. Under certain conditions, it can be shown that $\mathbb{P}$ is Markov $\mathcal{G}$ if and only if its probability function f can be written as

$$f(x) = \frac{\prod_{C \in \mathcal{C}} \psi_C(x_C)}{Z} \tag{18.1}$$

where $\mathcal{C}$ is the set of maximal cliques and

$$Z = \sum_x \prod_{C \in \mathcal{C}} \psi_C(x_C).$$

18.6 Example. The maximal cliques for the graph in Figure 18.1 are $C_1 = \{X, Y\}$ and $C_2 = \{Y, Z\}$. Hence, if $\mathbb{P}$ is Markov to the graph, then its probability function can be written

$$f(x, y, z) \propto \psi_1(x, y)\psi_2(y, z)$$

for some positive functions ψ_1 and ψ_2. ■

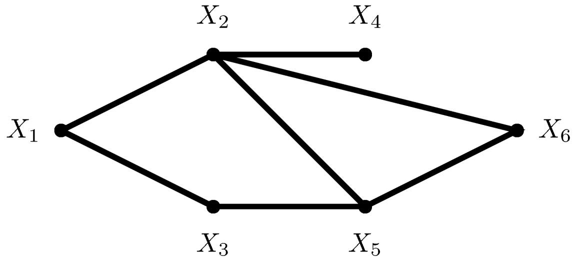

FIGURE 18.9. The maximumly cliques of this graph are $\{X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_4\}, \{X_3, X_5\}, \{X_2, X_5, X_6\}$.

18.7 Example. The maximal cliques for the graph in Figure 18.9 are

$$\{X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_4\}, \{X_3, X_5\}, \{X_2, X_5, X_6\}.$$

Thus we can write the probability function as

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5, x_6) &\propto \psi_{12}(x_1, x_2)\psi_{13}(x_1, x_3)\psi_{24}(x_2, x_4) \\ &\quad \times \psi_{35}(x_3, x_5)\psi_{256}(x_2, x_5, x_6). \quad \blacksquare \end{aligned}$$

18.4 Fitting Graphs to Data

Given a data set, how do we find a graphical model that fits the data? As with directed graphs, this is a big topic that we will not treat here. However, in the discrete case, one way to fit a graph to data is to use a **log-linear model**, which is the subject of the next chapter.

18.5 Bibliographic Remarks

Thorough treatments of undirected graphs can be found in Whittaker (1990) and Lauritzen (1996). Some of the exercises below are from Whittaker (1990).

18.6 Exercises

1. Consider random variables (X_1, X_2, X_3). In each of the following cases, draw a graph that has the given independence relations.

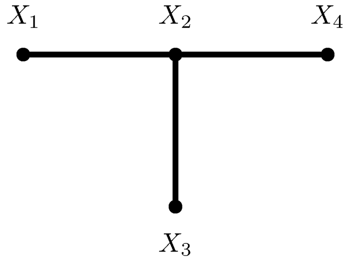

FIGURE 18.10.

FIGURE 18.11.

(a) $X_1 \amalg X_3 \mid X_2$.

(b) $X_1 \amalg X_2 \mid X_3$ and $X_1 \amalg X_3 \mid X_2$.

(c) $X_1 \amalg X_2 \mid X_3$ and $X_1 \amalg X_3 \mid X_2$ and $X_2 \amalg X_3 \mid X_1$.

2. Consider random variables (X_1, X_2, X_3, X_4). In each of the following cases, draw a graph that has the given independence relations.

   (a) $X_1 \amalg X_3 \mid X_2, X_4$ and $X_1 \amalg X_4 \mid X_2, X_3$ and $X_2 \amalg X_4 \mid X_1, X_3$.

   (b) $X_1 \amalg X_2 \mid X_3, X_4$ and $X_1 \amalg X_3 \mid X_2, X_4$ and $X_2 \amalg X_3 \mid X_1, X_4$.

   (c) $X_1 \amalg X_3 \mid X_2, X_4$ and $X_2 \amalg X_4 \mid X_1, X_3$.

3. A conditional independence between a pair of variables is **minimal** if it is not possible to use the Separation Theorem to eliminate any variable from the conditioning set, i.e. from the right hand side of the bar Whittaker (1990). Write down the minimal conditional independencies from: (a) Figure 18.10; (b) Figure 18.11; (c) Figure 18.12; (d) Figure 18.13.

4. Let X_1, X_2, X_3 be binary random variables. Construct the likelihood ratio test for

$$H_0 : X_1 \amalg X_2 | X_3 \quad \text{versus} \quad H_1 : X_1 \text{is not independent of } X_2 | X_3.$$

5. Here are breast cancer data from Morrison et al. (1973) on diagnostic center (X_1), nuclear grade (X_2), and survival (X_3):

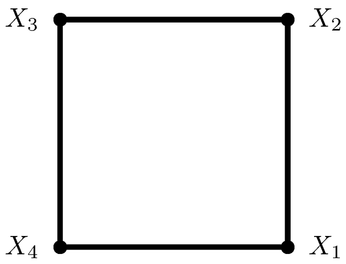

FIGURE 18.12.

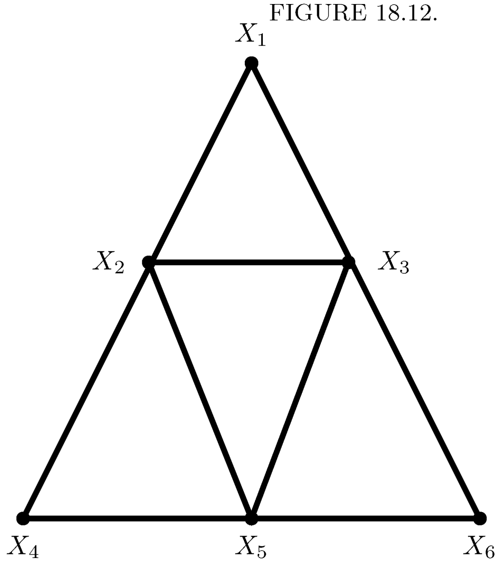

FIGURE 18.13.

	X_2	malignant	malignant	benign	benign
	X_3	died	survived	died	survived
X_1	Boston	35	59	47	112
	Glamorgan	42	77	26	76

(a) Treat this as a multinomial and find the maximum likelihood estimator.

(b) If someone has a tumor classified as benign at the Glamorgan clinic, what is the estimated probability that they will die? Find the standard error for this estimate.

(c) Test the following hypotheses:

$$
\begin{array}{lcl}
X_1 \amalg X_2|X_3 & \text{versus} & X_1 \leftrightsquigarrow X_2|X_3 \\
X_1 \amalg X_3|X_2 & \text{versus} & X_1 \leftrightsquigarrow X_3|X_2 \\
X_2 \amalg X_3|X_1 & \text{versus} & X_2 \leftrightsquigarrow X_3|X_1
\end{array}
$$

Use the test from question 4. Based on the results of your tests, draw and interpret the resulting graph.