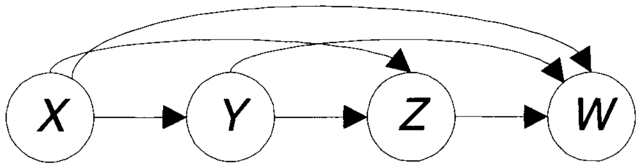

Figure 4.7: Every probability distribution of X, Y, Z, and W satisfies the Markov condition with this complete DAG.

Case	Sex	Height (inches)	Wage ($)
1	*female*	64	30,000
2	*female*	64	30,000
3	*female*	64	40,000
4	*female*	64	40,000
5	*female*	68	30,000
6	*female*	68	40,000
7	*male*	64	40,000
8	*male*	64	50,000
9	*male*	68	40,000
10	*male*	68	50,000
11	*male*	70	40,000
12	*male*	70	50,000

Table 4.1: Data on 12 workers.

4.2 Learning Structure (Model Selection)

Structure learning consists of learning the DAG in a Bayesian network from data. We want to learn a DAG that satisfies the Markov condition with the probability distribution P that is generating the data. Note that we do not know P; all we know are the data. The process of learning such a DAG is called **model selection**.

Example 4.11 *Suppose we want to model the probability distribution P of sex, height, and wage for American workers. We may obtain the data on* 12 *workers appearing in Table 4.1. We don't know the probability distribution P. However, from these data we want to learn a DAG that is likely to satisfy the Markov condition with P.*

If our only goal was simply learning some DAG that satisfied the Markov condition with P, our task would be trivial because a probability distribution P satisfies the Markov condition with every complete DAG containing the variables over which P is defined. We illustrate this with a DAG containing four variables. First, recall that a complete DAG is a DAG that has an edge between every pair of variables. Next consider the complete DAG in Figure 4.7. Each