5 Directed acyclic graphs

(5.1) Introduction In many statistical studies we have prior knowledge about a temporal or causal ordering of the variables. In this chapter we will use directed graphs to incorporate such knowledge into a graphical model for the variables.

Let $X_V = (X_v)_{v \in V}$ be a vector of real-valued random variables with probability distribution P and density p. Then the density p can always be decomposed into a product of conditional densities,

$$p(x) = p(x_d|x_1, \ldots, x_{d-1})p(x_1, \ldots, x_{d-1}) = \ldots = \prod_{v=1}^{d} p(x_v|x_1, \ldots, x_{v-1}). \tag{5.1}$$

Note that this can be achieved for any ordering of the variables. Now suppose that the conditional density of some variable X_v does not depend on all its predecessors, namely $X_1, \ldots, X_{v-1}$, but only on a subset X_{U_v}, that is, X_v is conditionally independent of its predecessors given X_{U_v}. Substituting $p(x_v|x_{U_v})$ for $p(x_v|x_1, \ldots, x_{v-1})$ in the product (5.1), we obtain

$$p(x) = \prod_{v=1}^{d} p(x_v|x_{U_v}). \tag{5.2}$$

This recursive dependence structure can be represented by a directed graph G by drawing an arrow from each vertex in U_v to v. As an immediate consequence of the recursive factorization, the resulting graph is acyclic, that is, it does not contain any loops.

On the other hand, P factorizes with respect to the undirected graph G^{m} which is given by the class of complete subsets $\mathscr{D} = \{\{v\} \cup U_v | v \in V\}$. This graph can be obtained from the directed graph G by completing all sets $\{v\} \cup U_v$ and then converting all directed edges into undirected ones. The graph G^{m} is called the moral graph of G since it is obtained by "marrying all parents of a joint child".

As an example, suppose that we want to describe the distribution of a genetic phenotype (such as blood group) in a family. In general, we can assume that the phenotype of the father (X) and that of the mother (Y) are independent, whereas the phenotype of the child (Z) depend on the phenotypes of both parents. Thus the joint density can be written as

$$p(x, y, z) = p(z|x, y)p(y)p(x).$$

The dependencies are represented by the directed graph G in Figure 5.1 (a). The absence of an edge between X and Y indicates that these variables are independent.

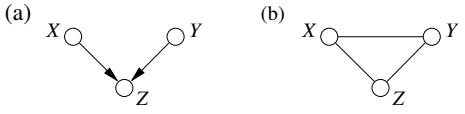

Figure 5.1: (a) Directed graph G representing the dependences between some phenotype of father (X), mother (Y), and child (Z); (b) moral graph G^{m} (b).

On the other hand, if we know the phenotype of the child then X and Y are no longer dependent. In the case of blood groups, for example, if the child has blood group AB and the father's blood group is A, then the blood group of the mother must be either B or AB. This conditional dependence of X and Y given Z is reflected by an edge in the moral graph G^{m}, which is complete and thus does not encode any conditional independence relations among the variables.

(5.2) Remark If the joint probability distribution has no density, we can still use conditional independences of the type

$$X_v \perp\!\!\!\perp X_1, \ldots, X_{v-1} \mid X_{U_j}$$

to associate a directed graph with P.

(5.3) Example (Markov chain) Consider a Markov chain on a discrete state space, i.e. a sequence $(X_t)_{t\in\mathbb{N}}$ of random variables such that

$$\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}).$$

Then the joint probability of $X_1, \ldots, X_T$ is given by

$$\mathbb{P}(X_T = x_T, \ldots, X_1 = x_1) = \prod_{t=2}^{T} \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) \cdot \mathbb{P}(X_1 = x_1).$$

The corresponding graph is depicted in Figure 5.3.

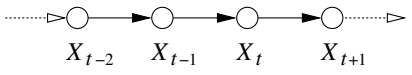

Figure 5.2: Directed graph G for a Markov chain X_t.

(5.4) Example (Regression) Let X_1, X_2, ε be independent $\mathcal{N}(0, \sigma^2)$ distributed random variables and suppose that

$$Y = \alpha X_1 + \beta X_2 + \varepsilon.$$

Since X_1 and X_2 are independent the joint density can be written as

$$p(x_1, x_2, y) = p(y|x_1, x_2)p(x_1)p(x_2).$$

On the other hand we have

$$\mathrm{cov}(X_1, X_2|Y) = \mathrm{cov}(X_1, X_2) - \mathrm{cov}(X_1, Y)\mathrm{var}(Y)^{-1}\mathrm{cov}(Y, X_2) = \frac{\alpha\beta}{1+\alpha^2+\beta^2}\sigma^2$$

which implies together with $\mathrm{cov}(X, Z|Y) = \mathrm{cov}(X, Z)$

$$p(x_1, x_2, y) = p(x_1|x_2, y)p(x_2|y)p(y).$$

Thus different orderings of the variables can lead to different directed graphs.

(5.5) Example As a last example, suppose that in a sociological study examining the causes for differences in social status the following four variables have been observed:

- Social status of individual's parents (S_p),
- Gender of individual (G),
- School education of individual (E),
- Social status of individual (S_i).

We assume that gender G and social status of the parents are independent reflecting the fact that gender is determined genetically and no selection has been taken place (this might not be true in certain cultures). Thus the joint density can be factorized recursively with respect to the graph in Figure 5.5.

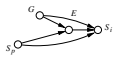

Figure 5.3: Directed graph G for sociological study about causes of social status of individuals.

The presence or absence of other edges reflects research hypotheses we might be interested in, as for example

- $G \longrightarrow E$: Is there a gender specific discrimination in education and does this depend on social class ($E \perp\!\!\!\perp G \mid S_p$)?
- $S_p \longrightarrow E$: Is there a social class effect in education and does this possibly depend on gender ($E \perp\!\!\!\perp S_p \mid G$)?
- $G \longrightarrow S_i$: Is there a gender specific discrimination in society - besides effects of education or social class of parents ($S_i \perp\!\!\!\perp G \mid E, S_p$)?
- $S_p \longrightarrow S_i$: Is there a discrimination of lower social class people - besides effects of education and gender ($S_i \perp\!\!\!\perp S_p \mid G, E$)?

In discussing the association between social status of the parents and education it is clear that we do not want to adjust for the social status of the individual as this might create a selection bias (an individual with low social status but with parents of high social status is likely to have had a bad education). To avoid such selection biases the the independence structure should be modelled by a directed graph.

(5.6) Directed acyclic graphs Let V be a finite and nonempty set. Then a *directed graph* G over V is given by an ordered pair (V, E) where the elements in V represent the vertices of G and $E \subseteq \{a \longrightarrow b \mid a, b \in V, a \neq b\}$ are the edges of G. If there exists an ordering $v_1, \ldots, v_d$ of the vertices which is consistent with the graph G, that is $v_i \longrightarrow v_j \in E$ implies $i < j$, then G is called a *directed acyclic graph* (DAG). It is clear that in that case G does not contain any cycle, i.e. a path of the

form $v \longrightarrow \ldots \longrightarrow v$. The vertices $\mathrm{pr}(v_j) = \{v_1, \ldots, v_{j-1}\}$ are the predecessors of v_j. The ordering is not uniquely determined by G.

Let $a \longrightarrow b$ be an edge in G. The vertex a is called a *parent* of b and b is a *child* of a. For $v \in V$ we define

- $\mathrm{ch}(v) = \{u \in V | v \longrightarrow u \in E\}$, the set of children of v,
- $\mathrm{pa}(v) = \{u \in V | u \longrightarrow v \in E\}$, the set of parents of v,
- $\mathrm{an}(v) = \{u \in V | u \longrightarrow \ldots \longrightarrow v \in E\}$, the set of *ancestors* of v,
- $\mathrm{de}(v) = \{u \in V | v \longrightarrow \ldots \longrightarrow u \in E\}$, the set of *descendants* of v,
- $\mathrm{nd}(v) = V \backslash \mathrm{de}(v)$, the set of *nondescendants* of v,

For $A \subseteq V$ we define $\mathrm{pa}(A) = \cup_{a \in A} \mathrm{pa}(a) \backslash A$ and similarly $\mathrm{ch}(A)$, $\mathrm{an}(A)$, $\mathrm{de}(A)$, and $\mathrm{nd}(A)$. Furthermore, we say that A is ancestral if $\mathrm{an}(A) \subseteq A$ and define $\mathrm{An}(A) = A \cup \mathrm{an}(A)$ as the *ancestral set* of A.

(5.7) Definition Let P be a probability distribution on the sample space $\mathscr{X}_V$ and $G = (V, E)$ a directed acyclic graph.

(DF)P *factorizes recursively* with respect to G if P has a density p with respect to a product measure μ on $\mathscr{X}_V$ such that

$$p(x) = \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)})$$

(5.8) Relation to separation in undirected graphs Suppose that P factorizes recursively with respect to a directed acyclic graph $G = (V, E)$. The factorization can be rewritten as

$$p(x_V) = \prod_{v \in V} \phi_{A_v}(x)$$

with $A_v = \{v\} \cup \mathrm{pa}(v)$ and $\phi_{A_v}(x) = p(x_v | x_{\mathrm{pa}(v)})$. Therefore P also factorizes with respect to any undirected graph in which the sets A_v are complete. To illustrate this relation it is sufficient to look only at graphs with three vertices and two edges. The three possible graphs are shown in Figure 5.8. The first graph is associated with the factorization

$$p(x) = p(x_3|x_2)p(x_2|x_1)p(x_1) = g(x_3, x_2)h(x_2, x_1),$$

which implies that $X_3 \perp\!\!\!\perp X_1 \mid X_2$. Similarly the second graph corresponds to

$$p(x) = p(x_3|x_2)p(x_1|x_2)p(x_2) = g'(x_3, x_2)h'(x_2, x_1),$$

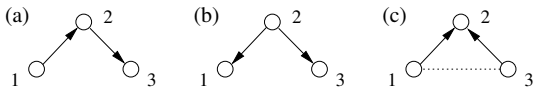

Figure 5.4: Basic configurations for three vertices and one missing edge.

which leads to the same conditional independence relation as above. The third graph, however, results from the factorization

$$p(x) = p(x_2|x_1, x_3)p(x_3)p(x_1).$$

Here, the first factor depends on all three variables and consequently X_1 and X_3 are no longer independent if we condition on X_2.

In general, two variables X_a and X_b become dependent when conditioning on a joint child X_c and therefore have to be joined ("married") in any undirected graph describing P. A subgraph induced by two non-adjacent vertices a and b and their common child c is called an immorality. Hence, a directed acyclic graph can be "moralized" by marrying all parents with a joint child.

(5.9) Definition Let $G = (V, E)$ be a directed acyclic graph. The *moral graph* of G is defined as the undirected graph $G^{\mathrm{m}} = (V, E^{\mathrm{m}})$ obtained from G by completing all immoralities in G and removing directions from the graph, i.e. for two vertices a and b with $a < b$ we have

$$a \text{ —— } b \in E^{\mathrm{m}} \Leftrightarrow a \longrightarrow b \in E \wedge \exists c \in V : \{a, b\} \in \mathrm{pa}(c).$$

(5.10) Proposition *Suppose P factorizes recursively with respect to G and A is an ancestral set. Then the marginal distribution P_A factorizes recursively with respect to the subgraph G_A induced by A.*

Proof. We have

$$p(x) = \prod_{v \in V} p(x_v|x_{\mathrm{pa}(v)}) = \prod_{v \in A} p(x_v|x_{\mathrm{pa}(v)}) \prod_{v \in V \setminus A} p(x_v|x_{\mathrm{pa}(v)}).$$

Since A is ancestral the first factor does not depend on $x_{V \setminus A}$ and the recursive facorization follows by integration over $x_{V \setminus A}$. □

(5.11) Corollary *Suppose P factorizes recursively with respect to G. Let A, B, and S be disjoint subsets of V. Then*

$$A \bowtie B \mid S \;\; [(G_{\mathrm{An}(A \cup B \cup S)})^{\mathrm{m}}] \;\Rightarrow\; X_A \perp\!\!\!\perp X_B \mid X_S.$$

Proof. Let $U = \mathrm{An}(A \cup B \cup S)$. By the previous proposition P_U factorizes recursively with respect to G_U which leads to

$$\begin{aligned} p(x_U) &= \prod_{v \in U} \underbrace{p(x_v|x_{\mathrm{pa}(v)})}_{\text{complete in } (G_U)^{\mathrm{m}}} \\ &= \prod_{\substack{A \subseteq U: \\ A \text{ complete}}} \psi_A(x_U), \end{aligned}$$

i.e. P_U factorizes with respect to $(G_U)^{\mathrm{m}}$. By Proposition 2.5 P satisfies the global Markov property with respect to $(G_U)^{\mathrm{m}}$ which implies that $X_A \perp\!\!\!\perp X_B \mid X_S$ whenever $A \bowtie B \mid S$ in $(G_U)^{\mathrm{m}}$. □

(5.12) Definition (Markov properties for DAGs) Let P be a probability distribution on the sample space $\mathscr{X}_V$ and $G = (V, E)$ a directed acyclic graph.

(DG)P satisfies the *global directed Markov property* with respect to G if for all disjoint sets $A, B, S \subseteq V$

$$A \bowtie B \mid S \;\; [(G_{\mathrm{an}(A \cup B \cup S)})^{\mathrm{m}}] \;\Rightarrow\; X_A \perp\!\!\!\perp X_B \mid X_S.$$

(DL)P satisfies the *local directed Markov property* with respect to G if

$$X_v \perp\!\!\!\perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)}.$$

(DO)P satisfies the *ordered directed Markov property* with respect to G if

$$X_v \perp\!\!\!\perp X_{\mathrm{pr}(v)} \mid X_{\mathrm{pa}(v)}.$$

(DP)P satisfies the *pairwise directed Markov property* with respect to G if for all $a, b \in V$ with $a \in \mathrm{pr}(b)$

$$a \longrightarrow b \notin E \;\Rightarrow\; X_a \perp\!\!\!\perp X_b \mid X_{\mathrm{nd}(b) \setminus a}.$$

(5.13) Theorem *Let P be a probability distribution on $\mathscr{X}_V$ with density p with respect to some product measure μ on $\mathscr{X}_V$. Then*

$$\textit{(DF)} \Leftrightarrow \textit{(DG)} \Leftrightarrow \textit{(DL)} \Leftrightarrow \textit{(DO)} \Rightarrow \textit{(DP)}.$$

(5.14) Remark If P has a positive and continuous density then for all properties are equivalent. The last three implications are still valid if P does not have a density.

Proof of Theorem 5.13. In Corollary 5.11 we have already shown that (DF) implies (DG). Since the set $\{v\} \cup \mathrm{nd}(v)$ is ancestral we have

$$\{v\} \bowtie \mathrm{nd}(v) \setminus \mathrm{pa}(v) \mid \mathrm{pa}(v) \;\; [(G_{\{v\} \cup \mathrm{nd}(v)})^{\mathrm{m}})]$$

and hence (DL) follows from (DG). Next, noting that $\mathrm{pr}(v) \subseteq \mathrm{nd}(v)$ we get

$$X_v \perp\!\!\!\perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)} \overset{\text{(C2)}}{\Rightarrow} X_v \perp\!\!\!\perp X_{\mathrm{pr}(v)} \mid X_{\mathrm{pa}(v)}.$$

which proves (DL) $\Rightarrow$ (DO). Factorizing p according to the ordering of the vertices we obtain with (DO)

$$\begin{aligned} p(x) &= \prod_{v \in V} p(x_v | x_{\mathrm{pr}(v)}) \\ &= \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)}), \end{aligned}$$

i.e. P satisfies (DF). Finally it suffices to show that (DL) implies (DP). Suppose that $a \in \mathrm{pr}(b)$ and $a \longrightarrow b \notin E$. Then $a \notin \mathrm{pa}(b)$ and $a \in \mathrm{nd}(b)$. Hence

$$X_v \perp\!\!\!\perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)} \;\Rightarrow\; X_a \perp\!\!\!\perp X_b \mid X_{\mathrm{nd}(b) \setminus a}$$

by application of (C2) and (C3). □

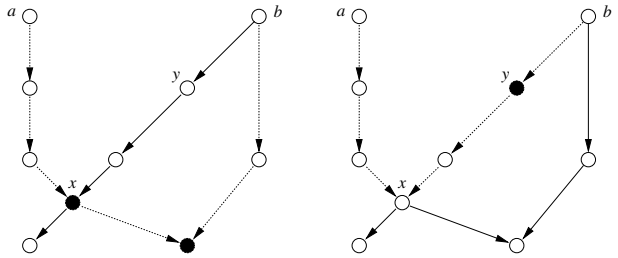

Figure 5.5: Illustration of the d-separation criterion: There are two paths between a and b (dashed lines) which are both blocked by $S = \{x, y\}$. For each paths the blocking is due to different intermediate vertices (filled circles).

(5.15) D-separation Let $\pi = \langle e_1, \ldots, e_n \rangle$ be a path between a and b with intermediate nodes $v_1, \ldots, v_{n-1}$. Then v_j is a *collider in* π if the adjecent edges in π meet head-to-head, i.e.

$$\langle e_j, e_{j+1} \rangle = v_{j-1} \longrightarrow v_j \longleftarrow v_{j+1}.$$

Otherwise we call v_j a *noncollider in* π.

Let S be a subset of V. The path π is *blocked by* S (or S-blocked) if it has an intermediate node v_j such that

- v_j is a collider and $v_j \notin \mathrm{An}(S)$ or
- v_j is a noncollider and $v_j \in S$.

A path which is not blocked by S is called S*-open*.

Let A, B, and S be disjoint subsets of V. Then S *d-separates* A and B if all paths between A and B are S-blocked. We write $A \bowtie_d B \mid S \ [G]$.

As an example consider the graph in Figure 5.15. It is sufficient to consider nonselfintersecting paths. There are two such paths between a and b, which are both blocked by $S = \{x, y\}$. The first path (left) is blocked by two vertices, a noncollider in S (x) and a collider which is not in S nor has any descendants in S. In the second path (right) y is a noncollider and therefore blocks the path.

(5.16) Theorem *Let $G = (V, E)$ be a directed acyclic graph. Then*

$$A \bowtie B \mid S \ [(G_{\mathrm{An}(A \cup B \cup S)})^{\mathrm{m}}] \Leftrightarrow A \bowtie_d B \mid S \ [G].$$

Proof. Suppose that π is an S-bypassing path from a to b in $(G_{\mathrm{An}(A \cup B \cup S)})^{\mathrm{m}}$. Every edge $a \text{ — } b$ in π such that a and b are not adjacent in G is due to moralization of $G_{\mathrm{An}(A \cup B \cup S)}$ and hence a and b have a common child c in $\mathrm{An}(A \cup B \cup S)$. Thus replacing $a \text{ — } b$ by $a \longrightarrow c \longleftarrow b$ (and converting all undirected edges into directed edges in G) we obtain a path π' in G which connects a and b and which does not have any noncolliders in S (since π was assumed to be S-bypassing).

Let $c_1 \ldots, c_r$ be the colliders in π' which have no descendants in S (i.e. $c_j \in \mathrm{An}(S)$) and which therefore block π'. We show by induction on r that an S-open path can be constructed from π. For $r = 0$ the path is already S-open. Now suppose

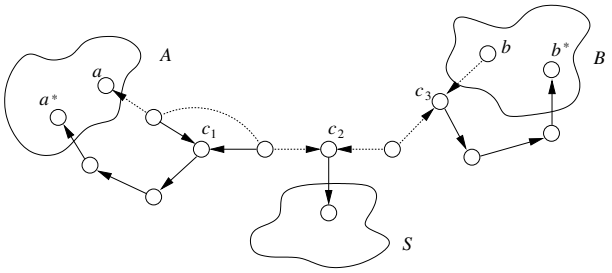

Figure 5.6: Construction of a S-open path between A and B from an S-bypassing path (dashed lines) in $(G_{\mathrm{An}(A\cup B\cup S)})^{\mathrm{m}}$.

that π' has r blocking colliders and that for paths with $r-1$ blocking colliders an S-open path can be constructed.

- If $c_1 \in \mathrm{An}(A)$ then there exists a directed path ϕ from c_1 to $a^* \in A$ which bypasses S (since $c_1 \notin \mathrm{An}(S)$). Let $\bar{\phi}$ be the reverse path from a to c_j and $\pi' = \langle \pi'_1, \pi'_2 \rangle$ where π'_1 is a path from a to c_1. It follows that $\pi'' = \langle \bar{\phi}, \pi'_2 \rangle$ is a path from a^* to b with $r-1$ blocking colliders. By assumption we can construct an S-open path between A and B from π''.
- If $c_1 \in \mathrm{An}(B)$ then there exists a directed path ϕ from c_1 to $b^* \in B$ which bypasses S. Let $\pi' = \langle \pi'_1, \pi'_2 \rangle$ where π'_1 is a path from a to c_1. It follows that $\pi'' = \langle \pi'_2, \phi \rangle$ is a path from a to b^* with $r-1$ blocking colliders. By assumption we can construct an S-open path between A and B from π''.

The construction of an S-open path between A and B from an S-bypassing path in $(G_{\mathrm{An}(A\cup B\cup S)})^{\mathrm{m}}$ is illustrated in Figure 5.

Conversely let π be an S-open path from A to B in G. Then the set C of all colliders in π is a subset of $\mathrm{An}(S)$ and hence π is also a path in $G_{\mathrm{An}(A\cup B\cup S)}$. Let $v_1, \ldots, v_{n-1}$ be the intermediate nodes of π. If $v_j \in S$ then $v_j \in C$ and thus $v_{j-1} \multimap v_{j+1} \in (E_{\mathrm{An}(A\cup B\cup S)})^{\mathrm{m}}$ with $v_{j-1}, v_{j+1} \notin S$. Leaving out all colliders in the sequence $v_1, \ldots, v_{n-1}$ and connecting consecutive nodes we therefore obtain an S-bypassing path in $(G_{\mathrm{An}(A\cup B\cup S)})^{\mathrm{m}}$. □

(5.17) Markov equivalence As we have illustrated in Example 5.4 the ordering of the variables can have an effect on the structure of the directed acyclic graph G representing P. On the other hand, two joint probability of two dependent variables X_a and X_b can be factorized either way leading to $a \longrightarrow b$ or $a \longleftarrow b$. Thus the question arises when two DAGs G_1 and G_2 do encode the same set of conditional independence relations?

(5.18) Theorem *Let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be two directed acyclic graphs over V. If G_1 and G_2 have*

- *the same skeleton (i.e. a and b are adjacent in G_1 if and only if they are adjacent in G_2) and*
- *the same immoralities (i.e. induced subgraphs of the form $a \longrightarrow c \longleftarrow b$)*

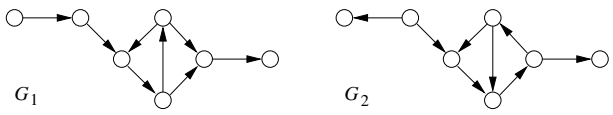

Figure 5.7: Two Markov equivalent graphs G_1 and G_2.

then they are Markov equivalent, that is, for all pairwise disjoint subsets A, B, and S of V

$$A \bowtie_d B \mid S \ \ [G_1] \ \Leftrightarrow \ A \bowtie_d B \mid S \ \ [G_2].$$

As an Example consider the graphs in Figure 5.17. Since both graphs G_1 and G_2 have the same skeleton and the same immorality they are Markov equivalent.

5.1 Recursive graphical models for discrete data

Let $X_V = (X_v)_{v \in V}$ be a vector of discrete valued random variables. A *recursive graphical model* for X_V is given by a directed acyclic grpah $G = (V, E)$ and a set $\mathscr{P}(G)$ of distributions that obey the Markov property with respect to G and thus factorize as

$$p(x) = \prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)}).$$

Using this factorization we can rewrite the likelihood function $\mathscr{L}(p)$ of X_V as

$$\begin{aligned}
\mathscr{L}(p) &\sim \prod_{i \in \mathscr{I}_V} \Big[\prod_{v \in V} p(x_v | x_{\mathrm{pa}(v)}) \Big]^{n(i)} \\
&= \prod_{v \in V} \prod_{i_{\mathrm{cl}(v)} \in \mathscr{I}_{\mathrm{cl}(v)}} \prod_{j \in \mathscr{I}_V : j_{\mathrm{cl}(v)} = i_{\mathrm{cl}(v)}} p(x_v | x_{\mathrm{pa}(v)})^{n(i)} \\
&= \prod_{v \in V} \prod_{i_{\mathrm{cl}(v)} \in \mathscr{I}_{\mathrm{cl}(v)}} p(x_v | x_{\mathrm{pa}(v)})^{n(i_{\mathrm{cl}(v)})} \\
&\sim \prod_{v \in V} \mathscr{L}_v(p),
\end{aligned}$$

where $\mathscr{L}_v(p)$ are the likelihood functions obtained when sampling the variables in $\mathrm{cl}(v)$ with fixed $\mathrm{pa}(v)$-marginals. It follows that the joint likelihood can be maximized by maximizing each factor $\mathscr{L}_v(p)$ separately.

Recall that for a complete graph G and given total count n the maximum likelihood estimate is given by $\hat{p}(i) = n(i)/n$. Similarly since the graph $G_{\mathrm{cl}(v)}$ is complete and the counts in the $\mathrm{pa}(v)$-marginal are fixed we have

$$\hat{p}(i_v | i_{\mathrm{pa}(v)}) = \frac{n(i_{\mathrm{cl}(v)})}{n(i_{\mathrm{pa}(v)})}.$$

(5.19) Theorem *The maximum likelihood estimator in the recursice graphical model for graph G is given by*

$$\hat{p}(i) = \prod_{v \in V} \frac{N(i_{\mathrm{cl}(v)})}{n(i_{\mathrm{pa}(v)})}.$$

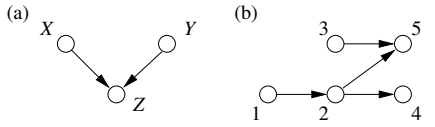

Figure 5.8: Two directed acyclic graphs G_1 and G_2.

(5.20) Example Suppose that G is of the form in Figure 5.1 (a). Then the MLE is of the form

$$\hat{p}_{ijk} = \frac{n_{ijk}}{n_{ij\cdot}} \frac{n_{\cdot j\cdot}}{n} \frac{n_{i\cdot\cdot}}{n}.$$

Despite the conditional independence of X and Y the data cannot be reduced beyond the table of counts n_{ijk} itself.

(5.21) Example Consider the graph G is Figure 5.1 (b). The corresponding MLE is given by

$$\hat{p}_{ijklm} = \frac{n_{i\cdots}}{n} \frac{n_{ij\cdots}}{n_{i\cdots}} \frac{n_{\cdot\cdot k\cdot\cdot}}{n} \frac{n_{\cdot j\cdot l\cdot}}{n_{\cdot j\cdots}} \frac{n_{\cdot jk\cdot m}}{n_{\cdot jk\cdot\cdot}}.$$