Stat 5421 Lecture Notes
Graphical Models
Charles J. Geyer
April 27, 2016

1 Introduction

Graphical models come in many kinds. There are graphical models where all the variables are categorical (Lauritzen, 1996, Chapter 4). There are graphical models where the variables are jointly multivariate normal (Lauritzen, 1996, Chapter 5). There are graphical models where some of the variables are categorical and the rest are conditionally jointly multivariate normal given the categorical ones (Lauritzen, 1996, Chapter 6).

Since this is a course in categorical data, we will only be interested in graphical models having only categorical variables.

2 Undirected Graphs

An *undirected graph* consists of a set whose elements are called *nodes* and a set whose elements are called *edges* and which are unordered pairs of nodes. We write $G = (N, E)$ where G is the graph, N is the node set, and E is the edge set.

Figure 1 (on p. 2) is an example. The nodes are U, V, W, X, and Z and the edges are the line segments.

A graph is called *simple* if it has no repeats of its edges, which our insistence that the edges are a set (which cannot have repeats) rather than a multiset (which can) already rules out, and if it has no loops (edges of the form (n, n)). We are only interested in simple graphs. The graph in Figure 1 is simple.

A graph (N_1, E_1) is a *subgraph* of (N_2, E_2) if $N_1 \subset E_1$ and $N_2 \subset E_2$, where $\subset$ means subset (we do not use $\subseteq$ to mean subset).

A graph is *complete* if every pair of nodes is connected by an edge. (Lauritzen, 1996, Section 2.1.1).

A *clique* in a graph is the node set of a maximal complete subgraph, one that is not a subgraph of another complete subgraph (Lauritzen, 1996, Section 2.1.1). The cliques in Figure 1 are $\{U\}$, $\{V, W\}$, and $\{W, X, Z\}$.

A *path* in an undirected graph is a sequence of edges (n_i, n_{i+1}), $i = 1$,

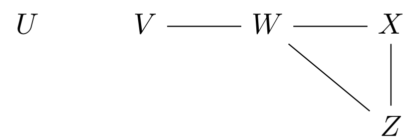

Figure 1: An Undirected Graph.

$\ldots$, k. For example,

$$V \text{ ------ } W \text{ ------ } X \text{ ------ } Z$$

is a path in the graph in Figure 1. We can say that this path goes from V to Z or from Z to V. Since the edges are undirected, there is no implied direction.

If A, B, and C are sets of nodes of a graph, then we say that C *separates* A and B if every path from a node in A to a node in B goes through C (has an edge, one of whose nodes is in C). In the graph in Figure 1, the node set $\{W\}$ separates the node sets $\{U, V\}$ and $\{X, Z\}$.

3 Undirected Graphs and Probability

A *log-linear model* for a contingency table is any model having one of the "sampling schemes" (Poisson, multinomial, product multinomial) described in Section 7 of the exponential families handout (Geyer, 2016).

Such a model is *hierarchical* if for every interaction term in the model all main effects and lower-order interactions involving variables in that term are also in the model.

The *interaction graph* for a hierarchical model is an undirected graph that has nodes that are the variables in the model and an edge for every pair of variables that appear in the same interaction term (Lauritzen, 1996, Section 4.3.3).

A hierarchical model is *graphical* if its terms correspond to the cliques of its interaction graph (Lauritzen, 1996, Section 4.3.3). Repeating what we said in the preceding section, the cliques the graph in Figure 1 are $\{U\}$, $\{V, W\}$, and $\{W, X, Z\}$. So the graphical model having this graph has formula

$$\sim U + V * W + W * X * Z.$$

Not every hierarchical model is graphical. For example, the hierarchical model with formula

$$\sim U + V * W + W * X + X * Z + Z * W$$

has the same interaction graph. But it is not graphical because it does not have the term $W * X * Z$.

A *Markov property* for a graph having random variables for nodes tells us something about conditional independence or factorization into marginals and conditionals. The relevant Markov property for log-linear models is the following, which is stated in Section 4.3.3 in Lauritzen (1996) and said to follow from Theorems 3.7 and 3.9 in Lauritzen (1996).

Theorem 1. *If C separates A and B in the interaction graph then the variables in A are conditionally independent of those in B given those in C.*

The conclusion of the theorem is often written

$$A \perp\!\!\!\perp B \mid C.$$

It means, of course, that the conditional distribution factorizes

$$f(y_{A \cup B} \mid y_C) = f(y_A \mid y_C) f(y_B \mid y_C)$$

where we have the usual abuse of notation that the three f's denote three different functions, the conditional PMF's of the indicated sets of variables, and where we are now using the notation y_S for subvectors of the response vector y of the probability model (described on p. 46 of the handout on exponential families, Geyer, 2016).

Repeating what we said in the preceding section, in the graph in Figure 1, $\{W\}$ separates $\{U, V\}$ and $\{X, Z\}$. So if Figure 1 is the interaction graph of a graphical model, we have

$$U, V \perp\!\!\!\perp X, Z \mid W.$$

Perhaps not quite so obvious, in the same graph the empty set separates U from all the other nodes. Since there are no paths from U to any other node. Every path from U to another node (there are none of them) goes through the empty set. Thus we can also say

$$U \perp\!\!\!\perp V, W, X, Z \mid \varnothing.$$

But, since conditioning on an empty set of variables is the same as not conditioning. We can also say

$$U \perp\!\!\!\perp V, W, X, Z$$

meaning U is (unconditionally) independent of the rest of the variables.

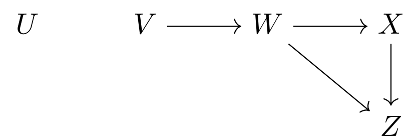

Figure 2: A Directed Graph.

4 Directed Graphs

A *directed graph* consists of a set whose elements are called *nodes* and a set whose elements are called *edges* and which are ordered pairs of nodes. When drawing a directed graph, the edges are drawn as arrows to indicate the direction. Figure 2 shows a directed graph.

A *path* in a directed graph is a sequence of edges (n_i, n_{i+1}), $i = 1, \ldots, k$. Since the edges are directed, there is an implied direction. We say the path goes from n_1 to n_{k+1}. For example,

$$V \longrightarrow W \longrightarrow X \longrightarrow Z$$

is a path from V to Z in the graph in Figure 2.

A directed graph is called *acyclic* if it has no cycles, which are paths that return to their starting point. The graph in Figure 2 is acyclic. Directed acyclic graphs are so important that they have a well-known TLA (three-letter acronym) DAG.

In a DAG the *parents* of a node u are the nodes from which edges go to u. The set of parents of u is denoted $\operatorname{pa}(u)$. In Figure 2 we have

$$\begin{aligned} \operatorname{pa}(U) &= \varnothing \\ \operatorname{pa}(V) &= \varnothing \\ \operatorname{pa}(W) &= \{V\} \\ \operatorname{pa}(X) &= \{W\} \\ \operatorname{pa}(Z) &= \{W, X\} \end{aligned}$$

5 Directed Acyclic Graphs and Probability

Every DAG that has nodes that are the variables in a statistical model is associated with a factorization of the joint distribution for that model into a product of marginal and conditional distributions

$$f(y) = \prod_{i \in N} f(y_i \mid y_{\mathrm{pa}(i)})$$

(Lauritzen, 1996, Section 4.5.1). Using the parent sets we found in the preceding section, the graph in Figure 2 has the factorization

$$f(u, v, w, x, z) = f(z \mid w, x) f(x \mid w) f(w \mid v) f(v) f(u). \tag{1}$$

This agrees with the separation properties of the corresponding undirected graph discussed in Section 3 above. A general factorization, valid for any probability model would be

$$\begin{aligned} f(u, v, w, x, z) &= f(z \mid u, v, w, x) f(u, v, w, x) \\ &= f(z \mid u, v, w, x) f(x \mid u, v, w) f(u, v, w) \\ &= f(z \mid u, v, w, x) f(x \mid u, v, w) f(w \mid u, v) f(u, v) \\ &= f(z \mid u, v, w, x) f(x \mid u, v, w) f(w \mid u, v) f(v \mid u) f(u) \end{aligned}$$

Comparing this general factorization with (1) we see that we have

$$\begin{aligned} f(z \mid u, v, w, x) &= f(z \mid w, x) \\ f(x \mid u, v, w) &= f(x \mid w) \\ f(w \mid u, v) &= f(w \mid v) \\ f(v \mid u) &= f(v) \end{aligned}$$

which imply

$$\begin{gathered} Z \perp\!\!\!\perp U, V \mid W, X \\ X \perp\!\!\!\perp U, V \mid W \\ W \perp\!\!\!\perp U \mid V \\ V \perp\!\!\!\perp U \end{gathered}$$

all of which are Markov properties that can be read off the undirected graph in Figure 1.

6 Directed Acyclic Graphs and Causality

DAGs are also widely used to indicate causal relationships (Pearl, 2009; Spirtes, et al., 2001). The idea is that a directed edge

$$V \longrightarrow W$$

indicates that changes in V cause changes in W.

Authorities on causal inference (Pearl, 2009; Spirtes, et al., 2001) understand that correlation is not causation. But they do insist that independence implies lack of causation.

If two variables are independent, then there cannot be a causal relationship between them (because that would induce dependence). If two variables are conditionally independent given a set S of other variables, then there cannot be a direct causal relationship. Any causal link must be indirect, the path from one to the other going through S.

Direction of causality cannot be inferred from independence, conditional or unconditional. If U and V are dependent, then changes in U may cause changes in V or changes in V may cause changes in U or changes in some other variable W may cause changes in both U and V.

There are two ways one can infer direction of causality. One is what is called "intervention" in the causal inference literature. It is the same thing that statisticians are talking about when they refer to controlled experiments. The other way is to simply assume that only one direction for edges makes sense (scientific sense, business sense, whatever).

In short, statistics can give you an undirected graph, but only controlled experiments or assumptions can give direction and causal interpretation of the edges.

References

Geyer, C. J. (2016). Exponential families, part I. Lecture notes for Stat 5421 (categorical data analysis). `http://www.stat.umn.edu/geyer/5421/notes/expfam.pdf`.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, New York.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*, 2nd ed. Cambridge University Press, Cambridge, UK.

Spirtes, P., Glymour, C., and Scheines, R. (2001). *Causation, Prediction, and Search*, 2nd ed. MIT Press, Cambridge, MA.