Independence Diagrams

Thomas P. Minka

Abstract

Independence diagrams are a graphical way of expressing the conditional independence relationships among a set of random variables. They cannot encode every possible form of conditional independence but they go a long way toward this end. They are also called "Bayesian networks," which unfortunately suggests inappropriate comparisons to neural networks. This paper discusses how to read and write independence diagrams. The presentation is based on Pearl (1988).

1 Introduction

The rules of probability theory require a certain amount of regularity in the conditional independence relationships among a set of random variables. For example, if A is independent of B given C, then B is independent of A given C. If A is independent of both B and D given C, then A must be independent of B given both C and D, etc.

Independence relationships are instrumental for simplifying calculations, but unfortunately they can be tedious to determine this way. Therefore it is a surprising and useful fact that the conditional independence relationships between a set of random variables can be elegantly expressed with a graph.

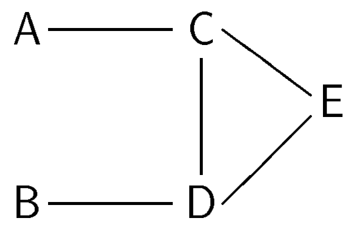
An independence diagram is a graph where each node represents a random variable. The conditional independence relationships between the variables are given by the graph-theoretic properties of *separation* and *d-separation*, hence can be immediately read off of the graph.

2 Undirected graphs

The simplest case is when all edges in the graph are undirected. The rule is:

> If there is no path from A to B, then variables A and B are independent.

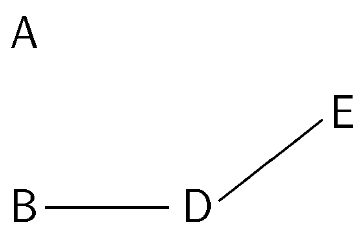
For example, in this graph:



A and B are dependent.

The converse of this rule is not true. If two variables are independent, then there may still be a path between them. No independence diagram can express all independence relationships perfectly, so we allow them to err conservatively. A path in the diagram therefore means "these variables could be dependent." A fully connected diagram is always "correct," in this sense, though not very useful. A graph is useful because of the edges that it does *not* have.

When the value of a variable is observed or otherwise known, then it is removed from the graph along with all edges connected to it. Since some paths may be broken, this expresses conditional independence. So if C is given, we get this graph:



where A and B are independent. Therefore, the original graph says that A and B are conditionally independent. Mathematically:

$$p(A|B,C) = p(A|C)$$

and

$$p(B|A,C) = p(B|C)$$

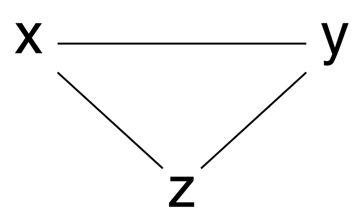
In the same graph, A and B are also independent given D alone, but they are not independent given E alone.

Another way to think about this rule is that when a variable is observed, it blocks the flow of information through it. Two variables are independent when all paths between them are blocked. If all paths between A and B are blocked by a set of variables $C_1..C_N$, then those variables are said to *separate* A and B in the graph. In this way, conditional independence between random variables maps onto separation in a graph. For example, you can check that the consistency properties mentioned in the introduction are always satisfied by separation in a graph.

The advantage of purely undirected independence diagrams is that the conditional independence relationships are simple. The disadvantage is that there is an important kind of conditional independence relationship that they cannot express, as discussed in the next section.

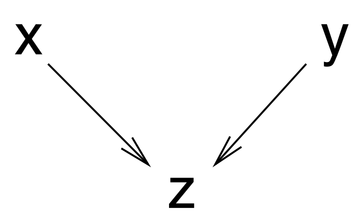
3 Directed graphs

Let $z = xy$, where random variables x and y are independent. The undirected independence diagram is



Why does there have to be an edge between x and y? Because if z is observed, then x and y are dependent (x must be z/y). Unfortunately, the resulting graph is not very interesting; it asserts no independence relationships at all.

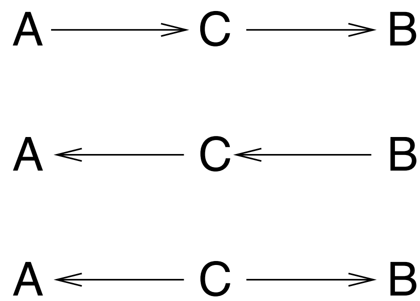
An undirected diagram is always a superset of the possible dependencies that can exist. So if a variable depends on the joint outcome of several other variables, then those other variables must be fully connected. What is missing is the ability for new dependencies to appear when we receive certain information. This is what directed edges will do for us. The directed independence diagram for this example is simply



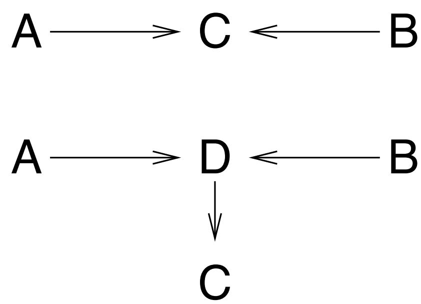
which asserts exactly the independence relationship we want: x is independent of y unless z is given.

The independence rules for directed graphs are best stated by example. In addition to the concept of a path being blocked by an observation, we now have to have the concept of a path being activated by an observation. A path may go against the arrows: the arrows only determine when a path should be blocked vs. activated.

Here are the cases where a path between A and B would be blocked by observing C (they would otherwise not be blocked):



and here are the cases where a path between A and B would be activated by observing C (they would otherwise be blocked):

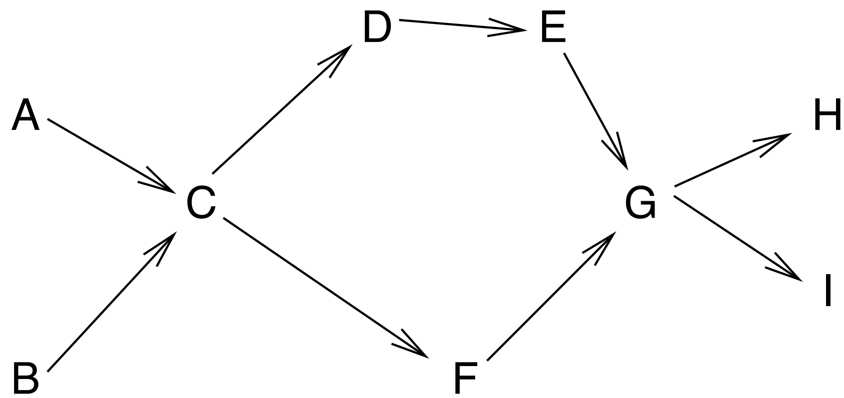


This is a more formal way to say it:

> A path between A and B is blocked if there is a node C such that
>
> 1. the path has converging arrows at C and none of C or its descendants are given, or
> 2. the path does not have converging arrows at C and C is given.
>
> If all paths between them are blocked, then A and B are independent. This kind of separation is called *d-separation*.

A descendant of C is any node on a directed path from C. The reason that we have to watch out for descendants of C is clear from the $z = xy$ example; suppose we have another random variable s which is the sign of z (thus making s a descendant of z). If we observe $s = -1$, then even if we don't know z, the variables x and y are dependent, since they must have different signs.

As a more detailed example, from this graph:



we can deduce that

- A is independent of B.
- E may be dependent on F.
- Given C, E is independent of F.
- Given both C and G, E may once again be dependent on F.
- Given C, H is independent of A.
- Given F, H may be dependent on A.
- Given I, A may be dependent on B.

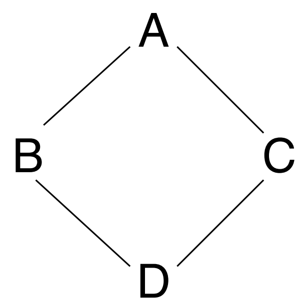
Mathematically, this diagram tells us that the joint distribution can be factored as:

$$p(A, B, C, D, E, F, G, H, I) = p(A)p(B)p(C|A, B)p(D|C)p(E|D)p(F|C)p(G|E, F)p(H|G)p(I|G)$$

This kind of factoring is true for any directed independence diagram with no directed cycles: each node is conditioned on the nodes pointing into it. The conditional independence statements listed above can also be deduced from this factoring. Thus directed graphs are simply another way to express a factoring of the joint distribution.

Often there are many directed graphs which encode the same conditional independence relationships. In this case, the arrows are usually chosen to reflect the flow of time or causality in the problem, which makes the conditional independence rules easier to remember (as in the $z = xy$ example). It also naturally avoids directed cycles. Of course, changing the arrows will change the implied factoring of the joint distribution.

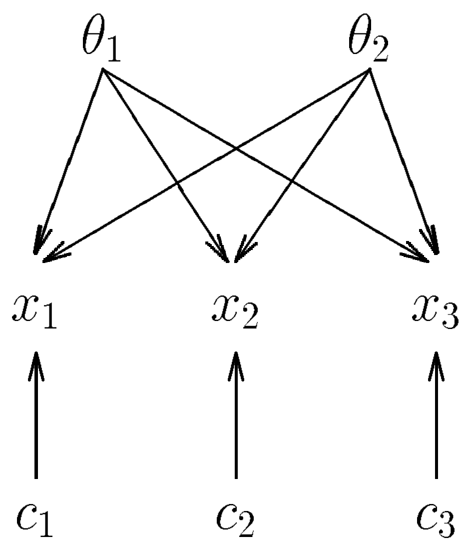
An advantage of purely directed independence diagrams is that they have a simple one-to-one relationship with factorings of the joint distribution. This makes them easy to read and easy to write. A disadvantage is that some conditional independence relationships cannot be expressed using directed edges alone. For example, this undirected graph:



has no equivalent directed graph (without a directed cycle), i.e. a graph that expresses exactly the same set of conditional independence relationships. Hence both directed and undirected edges are useful. It is possible to have both in the same graph, but the independence rules are involved so we will not discuss them here.

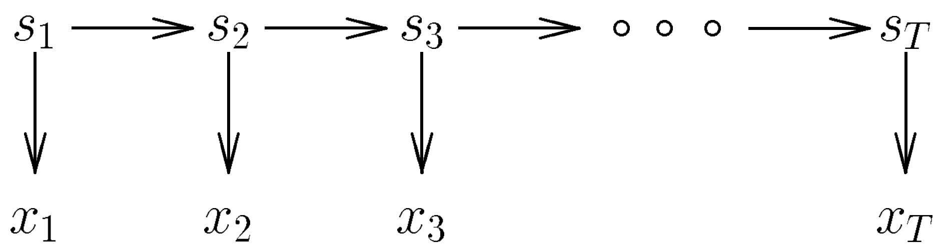
4 Examples

All of the statistical models we will consider in this course can be expressed naturally using independence diagrams. For example, a two-class recognition system can be expressed as:



where θ_1 and θ_2 are the parameters for each class, $x_1..x_3$ is the data to be classified, and $c_1..c_3$ are the classifications of the data. From the diagram, we can see that the data are independent once we know the class parameters, the class parameters are independent until we observe the data, etc.

The hidden Markov and linear dynamical system models we will do later both have a chain-type independence diagram:



The data is $x_1..x_T$ and the "hidden state" is $s_1..s_T$. The fact that the conditional independences have a chain structure will be the key to efficient probabilistic inference with these models.

5 Conclusion

The beauty of independence diagrams is that they impose minimal constraints on the joint distribution of variables, yet capture virtually everything you need to know to perform efficient inference. In particular, they tell you which variables should be relevant to any particular question.

Independence diagrams also tell you which variables, if known, would create the most independence between variables which are still unknown. Efficient algorithms for probabilistic computation work by exploiting these maximal cut-points in the independence graph.

Directed independence diagrams are also a convenient way to express a particular factoring of the joint distribution, and therefore express a parameterization of that distribution in terms of conditional probabilities. Thus there is great interest in Bayesian network tools which can infer the joint distribution of a data set, given prior knowledge in the form of an independence diagram.

Acknowledgements

I thank Rosalind Picard, Brian Clarkson, Tanzeem Choudhury, and Martin Szummer for improving the presentation.

References

[1] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Francisco, CA, 1988.