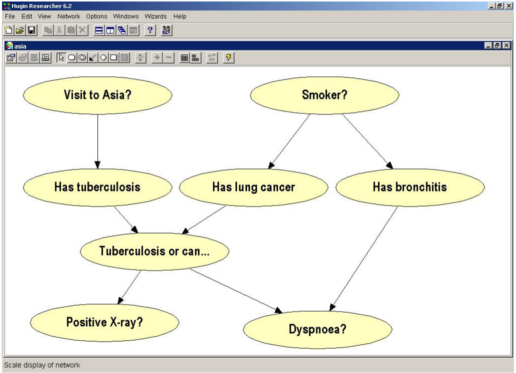
Graphs and Conditional Independence

Steffen Lauritzen, University of Oxford

Graphical Models, Lecture 1, Michaelmas Term 2011

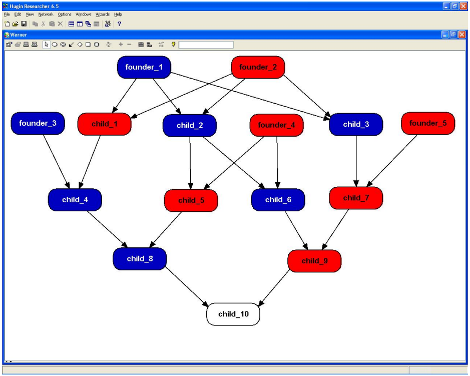
October 10, 2011

A directed graphical model



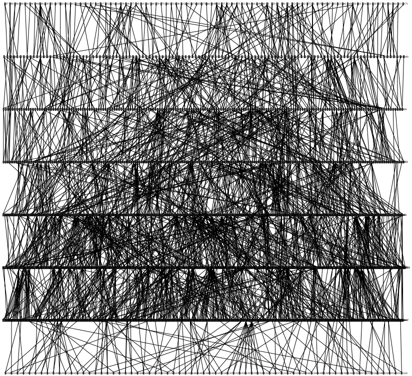
Directed graphical model (Bayesian network) showing relations between risk factors, diseases, and symptoms.

A pedigree



Graphical model for a pedigree from study of Werner's syndrome. Each node is itself a graphical model.

A large pedigree



Family relationship of 1641 members of Greenland Eskimo population.

Independence

We recall that two random variables X and Y are *independent* if

$$P(X \in A \mid Y = y) = P(X \in A)$$

or, equivalently, if

$$P\{(X \in A) \cap (Y \in B)\} = P(X \in A)P(Y \in B).$$

For discrete variables this is equivalent to

$$p_{ij} = p_{i+}p_{+j}$$

where $p_{ij} = P(X = i, Y = j)$ and $p_{i+} = \sum_j p_{ij}$ etc., whereas for continuous variables the requirement is a factorization of the joint density:

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

When X and Y are independent we write $X \perp\!\!\!\perp Y$.

Admissions to Berkeley by department

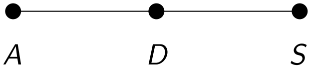
Here are three variables A: Admitted?, S: Sex, and D: Department.

Department	Sex	Whether admitted	
		Yes	No
I	Male	512	313
	Female	89	19
II	Male	353	207
	Female	17	8
III	Male	120	205
	Female	202	391
IV	Male	138	279
	Female	131	244
V	Male	53	138
	Female	94	299
VI	Male	22	351
	Female	24	317

When dealing with complex systems of many random variables, we must have a concept which is more sophisticated, but equally fundamental: that of *conditional independence*.

For three variables it is of interest to see whether independence holds for fixed value of one of them, e.g. *is the admission independent of sex for every department separately?*

We denote this as $A \perp\!\!\!\perp S \mid D$ and display it graphically as



Algebraically, this corresponds to the relations

$$p_{ijk} = p_{i+|k}\, p_{+j|k}\, p_{++k} = \frac{p_{i+k}\, p_{+jk}}{p_{++k}}.$$

Marginal and conditional independence

Note that there the two conditions

$$A \perp\!\!\!\perp S, \quad A \perp\!\!\!\perp S \mid D$$

are *very different* and will typically not both hold unless we either have $A \perp\!\!\!\perp (D, S)$ or $(A, D) \perp\!\!\!\perp S$, i.e. if one of the variables are completely independent of both of the others.

This fact is a simple form of what is known as *Yule–Simpson paradox.*

It can be much worse than this: A *positive conditional association can turn into a negative marginal association* and vice-versa.

Admissions revisited

Admissions to Berkeley

Sex	Whether admitted	
	Yes	No
Male	1198	1493
Female	557	1278

Note this marginal table shows much lower admission rates for females.

Considering the departments separately, there is only a difference for department I, and it is the other way around...

Admissions to Berkeley by department

Department	Sex	Whether admitted	
		Yes	No
I	Male	512	313
	Female	89	19
II	Male	353	207
	Female	17	8
III	Male	120	205
	Female	202	391
IV	Male	138	279
	Female	131	244
V	Male	53	138
	Female	94	299
VI	Male	22	351
	Female	24	317

Apart from Department I, it holds that $A \perp\!\!\!\perp S \mid D$. In Department I, a higher proportion of females are admitted!

Florida murderers

Sentences in 4863 murder cases in Florida over the six years 1973-78

	Sentence	
Murderer	Death	Other
Black	59	2547
White	72	2185

The table shows a greater proportion of white murderers receiving death sentence than black (3.2% vs. 2.3%), although the difference is not big, the picture seems clear.

Controlling for colour of victim

		Sentence	
Victim	Murderer	Death	Other
Black	Black	11	2309
	White	0	111
White	Black	48	238
	White	72	2074

Now the table for given colour of victim shows a very different picture. In particular, note that 111 white murderers killed black victims and none were sentenced to death.

Formal definition

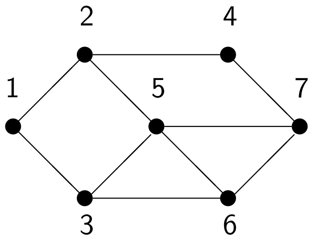
Random variables X and Y are *conditionally independent* given the random variable Z if

$$\mathcal{L}(X \mid Y, Z) = \mathcal{L}(X \mid Z).$$

We then write $X \perp\!\!\!\perp Y \mid Z$ (or $X \perp\!\!\!\perp_P Y \mid Z$)
Intuitively:
Knowing Z renders Y *irrelevant* for predicting X.
Factorisation of densities:

$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z \quad &\iff \quad f(x,y,z)f(z) = f(x,z)f(y,z) \\ &\iff \quad \exists a, b : f(x,y,z) = a(x,z)b(y,z). \end{aligned}$$

Undirected graphical models



For several variables, complex systems of conditional independence can for example be described by undirected graphs.

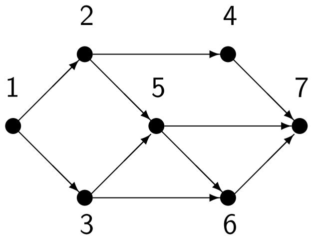
Then a set of variables A is conditionally independent of set B, given the values of a set of variables C if C separates A from B.

For example in picture above

$$1 \perp\!\!\!\perp \{4,7\} \,|\, \{2,3\}, \qquad \{1,2\} \perp\!\!\!\perp 7 \,|\, \{4,5,6\}.$$

Directed graphical models

Directed graphs are also natural models for conditional indpendence:



Any node is conditional independent of its non-descendants, given its immediate parents. So, for example, in the above picture we have

$$5 \perp\!\!\!\perp \{1,4\} \mid \{2,3\}, \quad 6 \perp\!\!\!\perp \{1,2,4\} \mid \{3,5\}.$$

For random variables X, Y, Z, and W it holds

(C1) If $X \perp\!\!\!\perp Y \mid Z$ then $Y \perp\!\!\!\perp X \mid Z$;

(C2) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp U \mid Z$;

(C3) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp Y \mid (Z, U)$;

(C4) If $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, then $X \perp\!\!\!\perp (Y, W) \mid Z$;

If density w.r.t. product measure $f(x, y, z, w) > 0$ also

(C5) If $X \perp\!\!\!\perp Y \mid (Z, W)$ and $X \perp\!\!\!\perp Z \mid (Y, W)$ then $X \perp\!\!\!\perp (Y, Z) \mid W$.

Proof of (C5)

We have

$$X \perp\!\!\!\perp Y \mid (Z, W) \Rightarrow f(x, y, z, w) = a(x, z, w)b(y, z, w).$$

Similarly

$$X \perp\!\!\!\perp Z \mid (Y, W) \Rightarrow f(x, y, z, w) = g(x, y, w)h(y, z, w).$$

If $f(x, y, z, w) > 0$ for all (x, y, z, w) it thus follows that

$$g(x, y, w) = a(x, z, w)b(y, z, w)/h(y, z, w).$$

The left-hand side does not depend on z so let $z = z_0$ be fixed. Then we have

$$g(x, y, w) = \tilde{a}(x, w)\tilde{b}(y, w).$$

Insert this into the second expression for f to get

$$f(x, y, z, w) = \tilde{a}(x, w)\tilde{b}(y, w)h(y, z, w) = a^*(x, w)b^*(y, z, w)$$

which shows $X \perp\!\!\!\perp (Y, Z) \mid W$.

Conditional independence can be seen as encoding abstract irrelevance. With the interpretation: *Knowing C, A is irrelevant for learning B,* (C1)–(C4) translate into:

(I1) If, knowing C, learning A is irrelevant for learning B, then B is irrelevant for learning A;

(I2) If, knowing C, learning A is irrelevant for learning B, then A is irrelevant for learning any part D of B;

(I3) If, knowing C, learning A is irrelevant for learning B, it remains irrelevant having learnt any part D of B;

(I4) If, knowing C, learning A is irrelevant for learning B and, having also learnt A, D remains irrelevant for learning B, then both of A and D are irrelevant for learning B.

The property analogous to (C5) is slightly more subtle and not generally obvious. Also the symmetry (C1) is a special property of *probabilistic conditional independence,* rather than of general irrelevance.

An *independence model* $\perp_\sigma$ is a ternary relation over subsets of a finite set V. It is *semi-graphoid* if for all subsets A, B, C, D:

(S1) if $A \perp_\sigma B \mid C$ then $B \perp_\sigma A \mid C$ (symmetry);

(S2) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ (decomposition);

(S3) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid (C \cup D)$ (weak union);

(S4) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid (B \cup C)$, then $A \perp_\sigma (B \cup D) \mid C$ (contraction).

It is a *graphoid* if (S1)–(S4) holds and

(S5) if $A \perp_\sigma B \mid (C \cup D)$ and $A \perp_\sigma C \mid (B \cup D)$ then $A \perp_\sigma (B \cup C) \mid D$ (intersection).

It is *compositional* if also

(S6) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ then $A \perp_\sigma (B \cup D) \mid C$ (composition).

Separation in undirected graphs

Let $\mathcal{G} = (V, E)$ be finite and simple undirected graph (no self-loops, no multiple edges).

For subsets A, B, S of V, let $A \perp_{\mathcal{G}} B \mid S$ denote that *S separates A from B in $\mathcal{G}$*, i.e. that all paths from A to B intersect S.

Fact: *The relation $\perp_{\mathcal{G}}$ on subsets of V is a compositional graphoid.*

This fact is the reason for choosing the name 'graphoid' for such independence model.

Systems of random variables

For a system V of *labeled random variables* $X_v, v \in V$, we use the shorthand

$$A \perp\!\!\!\perp B \mid C \iff X_A \perp\!\!\!\perp X_B \mid X_C,$$

where $X_A = (X_v, v \in A)$ denotes the variables with labels in A.

The properties (C1)–(C4) imply that $\perp\!\!\!\perp$ *satisfies the semi-graphoid axioms* for such a system, and the

An independence model of this kind is said to be *probabilistic*.

Probabilistic independence models are often not compositional.