

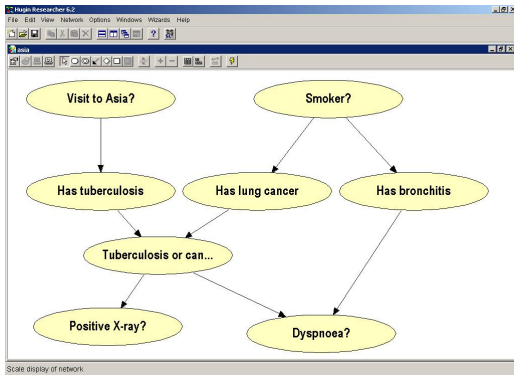
# Graphs and Conditional Independence

Steffen Lauritzen, University of Oxford

CIMPA Summerschool, Hammamet 2011, Tunisia

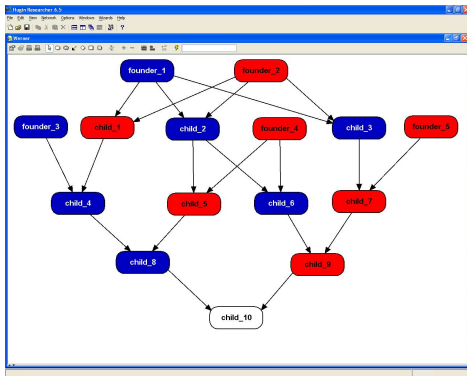
September 5, 2011

# A directed graphical model



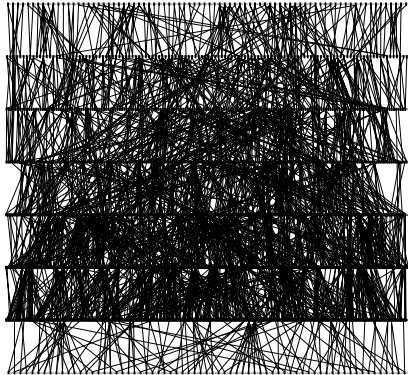
Directed graphical model (Bayesian network) showing relations between risk factors, diseases, and symptoms.

## 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.

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  $X \perp\!\!\!\perp Y$ .

## Formal definition

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

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

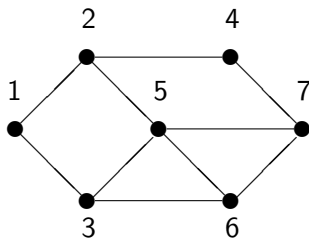
We then write  $X \perp\!\!\!\perp Y | Z$  (or  $X \perp\!\!\!\perp_P Y | Z$ )

Intuitively:

Knowing  $Z$  renders  $Y$  *irrelevant* for predicting  $X$ .

Factorisation of densities:

$$\begin{aligned} X \perp\!\!\!\perp Y | Z &\iff f_{XYZ}(x, y, z) f_Z(z) = f_{XZ}(x, z) f_{YZ}(y, z) \\ &\iff \exists a, b : f(x, y, z) = a(x, z) b(y, z). \end{aligned}$$



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

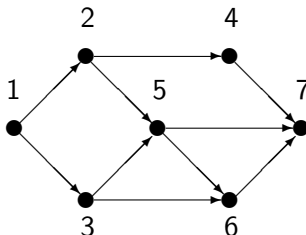
Then a set of variables  $A$  is conditionally independent of set  $B$ , given the values of a set of variables  $C$  if  $C$  separates  $A$  from  $B$ .

For example in picture above

$$1 \perp\!\!\!\perp \{4, 7\} \mid \{2, 3\}, \quad \{1, 2\} \perp\!\!\!\perp 7 \mid \{4, 5, 6\}.$$

## Directed graphical models

Directed graphs are also natural models for conditional independence:



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 for fixed  $z = z_0$ :

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

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.

An *independence model*  $\perp_\sigma$  is a ternary relation over subsets of a finite set  $V$ . It is *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*);
- (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*).

*Semigraphoid* if only (S1)–(S4) holds. 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 *graphoid axioms if the joint density of the variables is strictly positive*.

A regular *multivariate Gaussian distribution*, defines a *compositional graphoid independence model*.

## Geometric orthogonality

Let  $L$ ,  $M$ , and  $N$  be linear subspaces of a Hilbert space  $H$  and

$$L \perp M \mid N \iff (L \ominus N) \perp (M \ominus N),$$

where  $L \ominus N = L \cap N^\perp$ .  $L$  and  $M$  are said to *meet orthogonally in  $N$* .

- (O1) If  $L \perp M \mid N$  then  $M \perp L \mid N$ ;
- (O2) If  $L \perp M \mid N$  and  $U$  is a linear subspace of  $L$ , then  $U \perp M \mid N$ ;
- (O3) If  $L \perp M \mid N$  and  $U$  is a linear subspace of  $M$ , then  $L \perp M \mid (N + U)$ ;
- (O4) If  $L \perp M \mid N$  and  $L \perp R \mid (M + N)$ , then  $L \perp (M + R) \mid N$ .

*Intersection* does not hold in general whereas *composition* (S6) does.

$\mathcal{G} = (V, E)$  simple undirected graph; An independence model  $\perp_\sigma$  satisfies

(P) *the pairwise Markov property* if

$$\alpha \not\sim \beta \Rightarrow \alpha \perp_\sigma \beta \mid V \setminus \{\alpha, \beta\};$$

(L) *the local Markov property* if

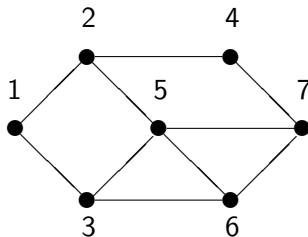
$$\forall \alpha \in V : \alpha \perp_\sigma V \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha);$$

(G) *the global Markov property* if

$$A \perp_{\mathcal{G}} B \mid S \Rightarrow A \perp_\sigma B \mid S.$$



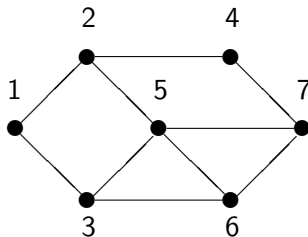
# Pairwise Markov property



Any non-adjacent pair of random variables are conditionally independent given the remaining.

For example,  $1 \perp_{\sigma} 5 \mid \{2, 3, 4, 6, 7\}$  and  $4 \perp_{\sigma} 6 \mid \{1, 2, 3, 5, 7\}$ .

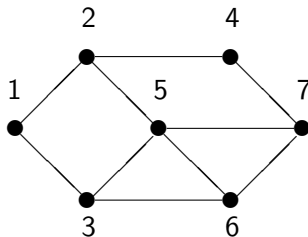
# Local Markov property



Every variable is conditionally independent of the remaining, given its neighbours.

For example,  $5 \perp_{\sigma} \{1, 4\} \mid \{2, 3, 6, 7\}$  and  $7 \perp_{\sigma} \{1, 2, 3\} \mid \{4, 5, 6\}$ .

# Global Markov property



To find conditional independence relations, one should look for separating sets, such as  $\{2, 3\}$ ,  $\{4, 5, 6\}$ , or  $\{2, 5, 6\}$

For example, it follows that  $1 \perp_{\sigma} 7 \mid \{2, 5, 6\}$  and  $2 \perp_{\sigma} 6 \mid \{3, 4, 5\}$ .

For any semigraphoid it holds that

$$(G) \Rightarrow (L) \Rightarrow (P)$$

If  $\perp_{\sigma}$  satisfies graphoid axioms it further holds that

$$(P) \Rightarrow (G)$$

so that *in the graphoid case*

$$(G) \iff (L) \iff (P).$$

The latter holds in particular for  $\perp\!\!\!\perp$ , when  $f(x) > 0$ .

$$(G) \Rightarrow (L) \Rightarrow (P)$$

(G) implies (L) because  $\text{bd}(\alpha)$  separates  $\alpha$  from  $V \setminus \text{cl}(\alpha)$ .

Assume (L). Then  $\beta \in V \setminus \text{cl}(\alpha)$  because  $\alpha \not\sim \beta$ . Thus

$$\text{bd}(\alpha) \cup ((V \setminus \text{cl}(\alpha)) \setminus \{\beta\}) = V \setminus \{\alpha, \beta\},$$

Hence by (L) and weak union (S3) we get that

$$\alpha \perp_{\sigma} (V \setminus \text{cl}(\alpha)) \mid V \setminus \{\alpha, \beta\}.$$

Decomposition (S2) then gives  $\alpha \perp_{\sigma} \beta \mid V \setminus \{\alpha, \beta\}$  which is (P).

$(P) \Rightarrow (G)$  for graphoids:

Assume  $(P)$  and  $A \perp_G B \mid S$ . *We must show  $A \perp_\sigma B \mid S$ .*

Wlog assume  $A$  and  $B$  non-empty. Proof is reverse induction on  $n = |S|$ .

If  $n = |V| - 2$  then  $A$  and  $B$  are singletons and  $(P)$  yields  $A \perp_\sigma B \mid S$  directly.

Assume  $|S| = n < |V| - 2$  and conclusion established for  $|S| > n$ :

First assume  $V = A \cup B \cup S$ . Then either  $A$  or  $B$  has at least two elements, say  $A$ . If  $\alpha \in A$  then  $B \perp_G (A \setminus \{\alpha\}) \mid (S \cup \{\alpha\})$  and also  $\alpha \perp_G B \mid (S \cup A \setminus \{\alpha\})$  (as  $\perp_G$  is a semi-graphoid). Thus by the induction hypothesis

$$(A \setminus \{\alpha\}) \perp_\sigma B \mid (S \cup \{\alpha\}) \text{ and } \{\alpha\} \perp_\sigma B \mid (S \cup A \setminus \{\alpha\}).$$

Now intersection (S5) gives  $A \perp_\sigma B \mid S$ .

## (P) $\Rightarrow$ (G) for graphoids, continued

For  $A \cup B \cup S \subset V$  we choose  $\alpha \in V \setminus (A \cup B \cup S)$ . Then  $A \perp_{\mathcal{G}} B \mid (S \cup \{\alpha\})$  and hence the induction hypothesis yields  $A \perp_{\sigma} B \mid (S \cup \{\alpha\})$ .

Further, either  $A \cup S$  separates  $B$  from  $\{\alpha\}$  or  $B \cup S$  separates  $A$  from  $\{\alpha\}$ . Assuming the former gives  $\alpha \perp_{\sigma} B \mid A \cup S$ .

Using intersection (S5) we get  $(A \cup \{\alpha\}) \perp_{\sigma} B \mid S$  and from decomposition (S2) we derive that  $A \perp_{\sigma} B \mid S$ .

The latter case is similar.