# Graphical and Log-Linear Models

## MSc Further Statistical Methods, Lecture 2
## Hilary Term 2006

Steffen Lauritzen, University of Oxford; January 23, 2006

# Conditional independence

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 $X \perp\!\!\!\perp Y \mid Z$

Intuitively:

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

Factorisation of probabilities:

$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z \quad &\iff \quad p_{xyz}p_{++z} = p_{x+z}p_{+y,z} \\ &\iff \quad \exists a, b : p_{xyz} = a_{yz}b_{yz}. \end{aligned}$$

# Fundamental properties

For any 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 $X \perp\!\!\!\perp U \mid Z$;

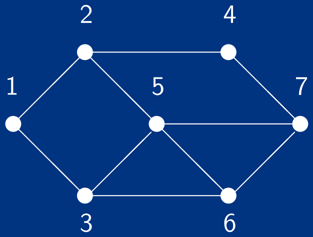
(C3) if $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp Y \mid (Z, U)$;

(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 all joint probabilities $p_{xyzw}$ are strictly positive also

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

# Graphical models




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

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

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

Algebraically the picture represents the fact that the joint probability of all variables factorizes into terms that only depends on *cliques* of the graph. In the pictures:

$$p_{ijklmno} = a_{ij} b_{ik} c_{jm} c_{jl} c_{kmn} c_{lo} c_{mno}.$$

Alternatively, the graph can be interpreted by saying that for each missing edge, there is a conditional independence associated:

$$I \not\sim J \implies I \perp\!\!\!\perp J \,|\, \text{remaining variables.}$$

# Formal Markov properties

Formally we say for a given graph $\mathcal{G}$ that a distribution obeys

(P) *the pairwise Markov property* if

$$\alpha \not\sim \beta \implies \alpha \perp\!\!\!\perp \beta \,|\, V \setminus \{\alpha, \beta\},$$

i.e. if all non-neighbours are conditionally independent given the remaining;

(L) *the local Markov property* if

$$\forall \alpha \in V : \alpha \perp\!\!\!\perp V \setminus \mathrm{cl}(\alpha) \,|\, \mathrm{bd}(\alpha),$$

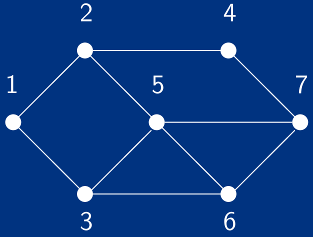
i.e. every variable is conditionally independent of the remaining given its neighbours;

(G) *the global Markov property* if

$$S \text{ separates } A \text{ from } B \text{ implies } A \perp\!\!\!\perp B \mid S.$$

*These are all equivalent if probabilities are strictly positive.*

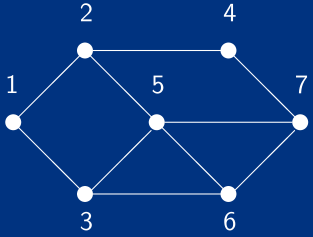
# Pairwise Markov property




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

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

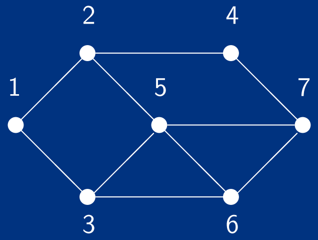
# Local Markov property




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

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

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

# Log–linear models

$\mathcal{A} = \{a_1, \ldots, a_K\}$ denotes a set of (pairwise incomparable) subsets of $a_i \subseteq V$.

A probability distribution $p$ (or function) *factorizes* w.r.t. $\mathcal{A}$ if it can be written as a product of terms where each only depend on variables in the same subset of $\mathcal{A}$.

The set of distributions which factorize w.r.t. $\mathcal{A}$ is the *log–linear model* generated by $\mathcal{A}$.

$\mathcal{A}$ is the *generating class* of the log–linear model.

It can be shown that a positive probability distribution is Markov w.r.t. a graph if and only if it factorizes as above with $a_i$ being *complete sets*, i.e. sets where all elements are mutual neighbours.

*If the distribution factorizes without being everywhere positive, it will also satisfy all the Markov properties,* but not the other way around.

Formally, we define the graphical model with graph $G = (V, E)$ to be the log-linear model with $\mathcal{A} = \mathcal{C}$, where $\mathcal{C}$ are the *cliques* (i.e. maximal complete subsets) of the graph.

# Example

Consider a three way contingency table, where e.g. $m_{ijk}$ denotes the mean of the counts $N_{ijk}$ in the cell $(i, j, k)$ which has then been expanded as e.g.

$$\log m_{ijk} = \alpha_i + \beta_j + \gamma_k \tag{1}$$

or

$$\log m_{ijk} = \alpha_{ij} + \beta_{jk} \tag{2}$$

or

$$\log m_{ijk} = \alpha_{ij} + \beta_{jk} + \gamma_{ik}, \tag{3}$$

or (with redundancy)

$$\log m_{ijk} = \gamma + \delta_i + \phi_j + \eta_k + \alpha_{ij} + \beta_{jk} + \gamma_{ik},$$

The additive terms in the expansion are known as *interaction terms of order* $|a| - 1$ or $|a|$*-factor interactions*.

Interaction terms of $0$th order are called *main effects*.

# Dependence graph

Any joint probability distribution of has a *dependence graph* $G = G(P) = (V, E(P))$.

This is defined by letting $\alpha \not\sim \beta$ in $G(P)$ exactly when

$$\alpha \perp\!\!\!\perp \beta \,|\, V \setminus \{\alpha, \beta\}.$$

$X$ will then satisfy the pairwise Markov w.r.t. $G(P)$ and the other Markov properties as well, in the case of positive probabilities.

# Dependence graph of log–linear model

For any generating class $\mathcal{A}$ we can construct the dependence graph of the corresponding log–linear model.

This is determined by the relation

$$\alpha \sim \beta \iff \exists a \in \mathcal{A} : \alpha, \beta \in a.$$

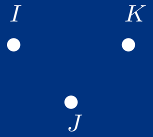
They are then also global, local, and pairwise Markov w.r.t. $G(\mathcal{A})$.

*This is by default the graph displayed in MIM.*

# Independence

The log–linear model specified by (1) is known as the *main effects model.*

It has generating class consisting of singletons only $\mathcal{A} = \{\{I\}, \{J\}, \{K\}\}$. It has dependence graph

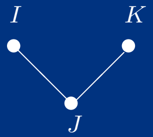



Thus it corresponds to *complete independence*.

# Conditional independence

The log–linear model specified by (2) has no interaction between $I$ and $K$.

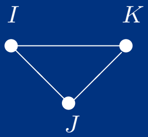
It has generating class $\mathcal{A} = \{\{I, J\}, \{J, K\}\}$ and dependence graph




Thus it corresponds to the *conditional independence* $I \perp\!\!\!\perp K \mid J$.

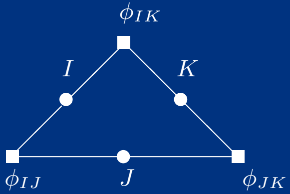
# No interaction of second order

The log–linear model specified by (3) has no second-order interaction. It has generating class $\mathcal{A} = \{\{I, J\}, \{J, K\}, \{I, K\}\}$ and its dependence graph




is the complete graph. Thus it has no conditional independence interpretation.

# Interaction graphs




The *interaction graph* of $\mathcal{A}$ is the graph with vertices $V \cup \mathcal{A}$ and edges define by

$$\alpha \sim a \iff \alpha \in a.$$

Using this graph all log–linear models admit a simple visual representation. Can be requested in MIM.

# Likelihood function

The likelihood function for an unknown $p$ can be expressed as

$$L(p) = \prod_{\nu=1}^{n} p(x^\nu) = \prod_{x \in \mathcal{X}} p(x)^{n(x)}.$$

In contingency table form the data follow a multinomial distribution

$$P\{N(x) = n(x), x \in \mathcal{X}\} = \frac{n!}{\prod_{x \in X} n(x)!} \prod_{x \in \mathcal{X}} p(x)^{n(x)}$$

but this only affects the likelihood function by a constant factor.

It can be shown that in log-linear models, *the likelihood function has at most one maximum*. When zero-values are allowed, it *always* has one.

MIM uses an algorithm for fitting known as *Iterative Proportional Fitting* which, if properly implemented, also works in the case where probabilities are allowed to be zero (sparse tables).

Also implemented e.g. in *R* in `loglin` with front end `loglm` in MASS.

An alternative is to "pretend" that counts are independent and Poisson distributed and use `glm`. However, the algorithm used there does *not* work when estimated cell probabilities are zero.