Log-linear models

Steffen Lauritzen, University of Oxford

Graphical Models, Lecture 3, Michaelmas Term 2010

October 18, 2010

A semi-graphoid relation $\perp_\sigma$ satisfies

(P) *the pairwise Markov property* if

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

(L) *the local Markov property* if

$$\forall \alpha \in V : \alpha \perp_\sigma V \setminus \mathrm{cl}(\alpha) \,|\, \mathrm{bd}(\alpha);$$

(G) *the global Markov property* if

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

It holds *for any semigraphoid* that (G) $\Rightarrow$ (L) $\Rightarrow$ (P) and *for a graphoid also* (P) $\Rightarrow$ (G)

Assume density f w.r.t. product measure on $\mathcal{X}$. For $a \subseteq V$, $\psi_a(x)$ denotes a function which depends on x_a only, i.e.

$$x_a = y_a \Rightarrow \psi_a(x) = \psi_a(y).$$

We can then write $\psi_a(x) = \psi_a(x_a)$ without ambiguity.

The distribution of X *factorizes w.r.t.* $\mathcal{G}$ or satisfies (F) if

$$f(x) = \prod_{a \in \mathcal{A}} \psi_a(x)$$

where $\mathcal{A}$ are *complete* subsets of $\mathcal{G}$.

Complete subsets of a graph are sets with all elements pairwise neighbours.

Let (F) denote the property that f factorizes w.r.t. $\mathcal{G}$ and let (G), (L) and (P) denote Markov properties w.r.t. $\perp\!\!\!\perp$. *It then holds that*

$$(F) \Rightarrow (G)$$

and further: *If $f(x) > 0$ for all x*, $(P) \Rightarrow (F)$.

Thus *in the case of positive density all Markov properties coincide:*

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

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

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

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

X will then satisfy the pairwise Markov w.r.t. $G(P)$ and $G(P)$ is smallest with this property, i.e. *P is pairwise Markov w.r.t. any graph $\mathcal{G}$ iff*

$$G(P) \subseteq \mathcal{G}.$$

If $f(x) > 0$ for all x, or P factorizes w.r.t. $G(P)$, *P is also globally Markov w.r.t. $G(P)$.*

Let $\mathcal{A}$ denote an arbitrary set of subsets of V. A density f (or function) *factorizes* w.r.t. $\mathcal{A}$ if there exist functions $\psi_a(x)$ which depend on x_a only and

$$f(x) = \prod_{a \in \mathcal{A}} \psi_a(x).$$

Similar to factorization w.r.t. graph, but $\mathcal{A}$ are not necessarily complete subsets of a graph.

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

To avoid redundancy, it is common to assume the sets in $\mathcal{A}$ to be incomparable in the sense that no subset in $\mathcal{A}$ is contained in any other member of $\mathcal{A}$. $\mathcal{A}$ is the *generating class* of the log–linear model.

Log-linear models are traditionally used for contingency tables, 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.

$$\log m_{ijk} = \alpha_i + \beta_j + \gamma_k \qquad (1)$$

or

$$\log m_{ijk} = \alpha_{ij} + \beta_{jk} \qquad (2)$$

or

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

or (with redundancy)

$$\log m_{ijk} = \gamma + \delta_i + \phi_j + \eta_k + \alpha_{ij} + \beta_{jk} + \gamma_{ik}, \qquad (4)$$

This largely a matter of different notation. Assume data $X^1 = x^1, \ldots, X^n = x^n$ and $V = \{I, J, K\}$ and write $i = 1, \ldots, |I|$ for the possible values of X_I etc. and

$$N_{ijk} = |\{\nu : x^\nu = (i, j, k)\}|,$$

etc. Then $m_{ijk} = nf(x)$ and *if $f(x) > 0$ and factorizes w.r.t. $\mathcal{A} = \{\{I, J\}, \{J, K\}\}$* we have

$$\log f(x) = \log \psi_{IJ}(x_I, x_J) + \log \psi_{JK}(x_J, x_K).$$

Thus if we let

$$\alpha_{ij} = \log n + \log \psi_{IJ}(x_I, x_J), \quad \beta_{jk} = \log \psi_{JK}(x_J, x_K)$$

we have

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

The only difference is the assumption of positivity which is not necessary when using the multiplicative definition.

The logarithm of the factors $\phi_a = \log \psi_a$ are known as *interaction terms of order* $|a| - 1$ or $|a|$*-factor interactions*.

Interaction terms of 0th order are called *main effects*.
We also refer to the factors themselves (rather than their logarithms) using the same terms.

Recall that a joint probability distribution P of $X = (X_v, v \in V)$ has a *dependence graph* $G = G(P) = (V, E(P))$, defined by letting $\alpha \not\sim \beta$ in $G(P)$ exactly when

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

The *dependence graph $G(\mathcal{P})$ for a family $\mathcal{P}$* is the smallest graph $\mathcal{G}$ so that all $P \in \mathcal{P}$ are pairwise Markov w.r.t. $\mathcal{G}$:

$$\alpha \perp\!\!\!\perp_P \beta \mid V \setminus \{\alpha, \beta\} \text{ for all } P \in \mathcal{P}.$$

For any generating class $\mathcal{A}$ we construct the dependence graph $G(\mathcal{A}) = G(\mathcal{P}_\mathcal{A})$ of the log–linear model $\mathcal{P}_\mathcal{A}$.

Since the pairwise Markov property has to hold for all members of $\mathcal{P}_\mathcal{A}$, it has at least to hold for all positive members. *The dependence graph is determined by the relation*

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

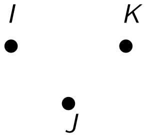
For sets in $\mathcal{A}$ are clearly complete in $G(\mathcal{A})$ and therefore *distributions in $\mathcal{P}_\mathcal{A}$ do factorize according to $G(\mathcal{A})$*. On the other hand, any graph with fewer edges would not suffice.

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

Independence

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

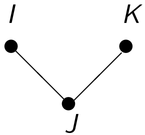
It has generating class consisting of singletons only $\mathcal{A} = \{\{I\}, \{J\}, \{K\}\}$. It has dependence graph



Thus it corresponds to *complete independence*.

Conditional independence

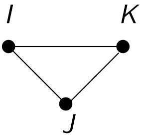
The log–linear model specified by (2) has no interaction between I and K.
It has generating class $\mathcal{A} = \{\{I, J\}, \{J, K\}\}$ and dependence graph



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.

As a generating class defines a dependence graph $G(\mathcal{A})$, the reverse is also true.

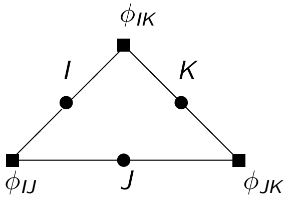
The set $\mathcal{C}(\mathcal{G})$ of *cliques* (maximal complete subsets) of $\mathcal{G}$ is a generating class for the log–linear model of distributions which factorize w.r.t. $\mathcal{G}$.

If the dependence graph completely summarizes the restrictions imposed by $\mathcal{A}$, i.e. if

$$\mathcal{A} = \mathcal{C}(G(\mathcal{A})),$$

$\mathcal{A}$ is *conformal*.

The generating classes for the models given by (1) and (2) are conformal, whereas this is not the case for (3).



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

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

Using this graph even non-conformal log–linear models admit a simple visual representation.

If $\mathcal{F} = F(\mathcal{A})$ is the factor graph for $\mathcal{A}$ and $\mathcal{G} = G(\mathcal{A})$ the corresponding dependence graph, it is not difficult to see that for A, B, S being subsets of V

$$A \perp_{\mathcal{G}} B \,|\, S \iff A \perp_{\mathcal{F}} B \,|\, S$$

and hence conditional independence properties can be read directly off the factor graph also.

In that sense, the factor graph is more informative than the dependence graph.

Note that David Edwards' program MIM, `www.hypergraph.dk`, uses the term *interaction graph* for the factor graph.