Forensic Statistics and Graphical Models (5)

Richard Gill
Spring Semester 2015
http://www.math.leidenuniv.nl/~gill/teaching/graphical

Graphical Models (DAGraphs)

- Directed acyclic graph (V, G)
- Vertex $v \in V$: random variable X_v
- Arrows $v \rightsquigarrow w \in G$: (direct) statistical (or causal) dependence
- $\mathrm{pa}(v)$: <u>graph parents</u> of vertex v
- For each vertex v, probability distributions $\mathrm{law}(X_v \mid X_{\mathrm{pa}(v)})$, represented by "conditional probability tables" $p(x_v \mid x_{\mathrm{pa}(v)})$
- Joint probability density: $p(x_{\{v\,:\,v \in V\}}) = \prod_{v \in V} p(x_v \mid x_{\mathrm{pa}(v)})$

Graphical Models (DAGraphs)

- <u>Ancestral</u> graph: include any ancestors

- <u>Moralized</u> graph: marry unmarried parents, forget direction of arrows

- **Theorem:** For any A, B, $C \subseteq V$,
  C separates A from B in mor(An($A \cup B \cup C$))
  $\Rightarrow X_A \perp X_B \mid X_C$

- **Theorem:** "C separates A from B in mor(An($A \cup B \cup C$))" is equivalent to <u>Pearl's d-separation criterium</u>

Pearl's d-separation criterium

- C <u>d-separates</u> A from B if all <u>trails</u> between A and B are <u>blocked</u> by C

- A <u>trail</u> is a path in the undirected graph obtained from G by forgetting the directions of the arrows

- A trail is <u>blocked</u> by C if it contains any one of the following:

    - a <u>chain</u> $u \rightarrow m \rightarrow v$ with middle node m in C

    - a <u>fork</u> $u \leftarrow m \rightarrow v$ with middle node in C

    - a <u>collider</u> $u \rightarrow m \leftarrow v$ with middle node not in C and moreover, none of the node's descendants in C

Pearl's d-separation criterium (reverse)

- *C* <u>d-connects</u> *A* and *B* if there exists a <u>trail</u> between *A* and *B* <u>opened</u> by *C*
- A trail is <u>opened</u> by *C* if all non-collider nodes on the trail are off *C* and any collider node is either in *C* or is among the ancestors of *C*
- *C* d-separates *A* from *B* if it doesn't d-connect them

Markov graph models (undirected graphs)

- <u>Undirected</u> graph (V, G)
- Random variables X_v , $v \in V$; joint density or mass function $p(x_{\{v\,:\,v \in V\}})$
- Notation: $X_C = X_{\{v\,:\,v \in C\}}$, $x_C = x_{\{v\,:\,v \in C\}}$; $X = X_V$, $x = x_V$
- For each complete subset C of the graph, a potential function ϕ_C
- $p(x) = \prod_C \phi_C(x_C)$
    - Wlog, “proportional to” is enough
    - Wlog, restrict C to cliques of G
- We say: “p factors over the graph G”

Markov graph models (undirected graphs)

- **Theorem:** Given, a Markov graph model
  Then, for any A, B, $C \subseteq V$,
  C separates A from B in $(G, V) \;\Rightarrow\; X_A \perp X_B \mid X_C$

- **Theorem:** (Clifford-Hammersley): Given, r.v.'s X_v, $v \in V$
  Suppose their joint density satisfies $p(x) > 0$ for all x
  Define a graph G by $v \nsim v'$ iff $X_v \perp X_{v'} \mid X_{V \setminus \{v, v'\}}$
  Then p factors over the graph G, i.e., the r.v.'s follow a Markov graph model w.r.t. G.

Chordal Markov graph models (undirected chordal graphs)

- Suppose $p(x) = \prod_C \phi_C(x_C)$ where C runs through cliques of G

- Suppose cliques can be arranged in a junction tree
  (G is triangulated, or chordal: no chordless cycles of length ≥ 4)

- Define <u>separator</u> S of two adjacent cliques on junction tree to be their intersection

- Initialise $\psi_S(x_S) = 1$ for all separators

- $p(x) = \prod_C \phi_C(x_C) / \prod_S \psi_S(x_S)$

Chordal Markov graph models: marginalisation

- For given $A \subseteq B$, and given ϕ_B, denote the <u>marginalisation</u> of ϕ_B over A by $\phi_{B\downarrow A}$
- If a <u>leaf</u> C of the junction tree contains vertices v not occurring elsewhere in the graph, then we can <u>marginalise</u> the corresponding variables out, by marginalising the potential ϕ_C over the separator S with the adjacent clique C', and absorbing both $\phi_{C\downarrow S}$ and ψ_S into $\phi_{C'}$

Markov graph models: conditioning

- We can condition out any of the variables of a Markov graph model, by redefining all potentials which include these variables to be zero off the instantiated values … and renormalising

- When conditioning *and* marginalising in some joint distribution, order of marginalisation and renormalisation is arbitrary!

- In other words, we may always postpone *renormalisation* to any convenient later moment

Chordal Markov graph models: the junction tree algorithm

- Consider two adjacent nodes C and C' of the junction tree, with separator S

- In a factorisation $p(x) \propto f(x) = \prod_C \phi_C(x_C) \,/\, \prod_S \psi_S(x_S)$ of joint density (or joint conditional density) along the junction tree, replace $\phi_C \, \psi_S^{-1} \phi_{C'}$ by $\phi_C \, (\phi_{C\downarrow S})^{-1} \, (\phi_{C\downarrow S} \, \psi_S^{-1} \phi_{C'})$

- Work from leaves <u>inwards</u> to any chosen "root" node, any way you like, then from root <u>outwards</u> to leaves, any way you like. (In effect: remove leaves, one by one, then put leaves back on, one by one)

- At completion of the "two passes", **every potential** (both those of cliques C and separators S) **contains the joint margin** of f, **for the corresponding variables** C or S