2

26/04/2020

DGMS

- undirected gms - Q $P(X_1, \ldots, X_8) = \frac{1}{Z} \exp\{E(X_1) + \ldots E(X_6, X_5, X_8)\}$ - Q

- DGMS → expert systems (complex causality webs)

HMM/dishonest casino

(*) key areas to focus on → 10708 learning objectives

· Areas to focus on:- - These should be included in 'lectures'

- BN definition
- Formal specification of factorisation
- Distinction between functional form of factorisation for DGM, UGMS
- Understand significance of conditional independence in PGMs
- Know the 3 local structures visually, what C.I relations they represent
- Nuances of I-Maps; appl. to example (*)
- Local Markovian assumptions
- d-separation: semantics, operation
· Global Markovian assumptions - Bayes ball.
· Equivalence theorem; difference with I-maps (*)
· Soundness + completeness in context of (*).

Jordan (2003) Ch 2 (read; annotate/supplement)

· very clear formal setup
· in general, not making use of additional 'structure', a joint probability distn of n random variables, each taking on r states requires full specification of r^n states in an n-dimensional probability table
· graphical models represent this more economically with ass. about structure via 'local relationships'
· formal notation:-
· Directed graph $G(V, E)$ V - set of nodes E - edges G - acyclic
· 1-to-1 mapping r.v.s to nodes
· For each node $i \in V$ → r.v. X_i
· $V = \{1, \ldots, n\}$ and r.v.s $= \{X_1, \ldots, X_n\}$

- Each node has a set of parent nodes, may be empty
- For each node $i \in V$, we denote π_i as the set of parents of node i
- The set of random variables X_{π_i} are the parents of the r.v. X_i
- Parent-child relations $\rightarrow$ locality $\rightarrow$ economical rep. of joint p.d.
- To each node i; associate function $f_i(x_i, x_{\pi_i})$ with properties:-

$$f_i(x_i, x_{\pi_i}) \geq 0 \quad \text{and} \quad \sum_{x_i} f(x_i, x_{\pi_i}) = 1$$

- The pre-conditional probability formalism is:
- Let $V = \{1, \ldots, n\}$; given a set of functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, define joint p.d:-

$$p(x_1, \ldots, x_n) \triangleq \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}) \qquad (2.1)$$

- Have to verify definition obeys joint p.d. constraints
- Choice of numerical values for $f_i \Rightarrow$ generation of specific joint p.d.
- Ranging over all numerical values $\Rightarrow$ family of joint prob distr. associated with graph G (*)
- This family can be characterised as products of local functions or graph-theoretically (edge patterns)

(*) The relationship between different ways of characterising the family of probability distributions associated with a graph that is key to PGM

- Conditional probability formalism:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_{\pi_i}) \qquad (2.2)$$

- $p(x_i \mid x_{\pi_i})$ - local condition prob. associated with graph G
- Building blocks whereby joint p.d. synthesised associated with G.
- Example given: (see diag).

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3)p(x_6 \mid x_2, x_5) \qquad (2.3)$$

- Product of LCD

Representational economy:

- r^n joint prob. table (no structure) $\longrightarrow$ r^{m_i+1} table (graph structure)
  - m_i - no. of parent nodes of X_i
  - local. cond. distr. X_i - (m_i+1) dim table
  - each node/r.v. has r values

(i.e.): exponential growth in n (no. variables in domain) exchanged for exponential growth in m_i (no. of parents of indiv. nodes X_i (fan-in))

- small fan-in $\rightarrow$ enormous reduction in complexity
- (i.e.) But there is more; not just data structure promise, but inferential machinery

---

2.1.1 Conditional independence

- independence:

x_A and x_B are independent, i.e. $X_A \perp\!\!\!\perp X_B$ if:-

$$p(x_A, x_B) = p(x_A)\,p(x_B) \qquad (2.4)$$

- conditional independence

X_A and X_C are conditionally independent, given X_B if:-

$$p(x_A, x_C \mid x_B) = p(x_A \mid x_B)\,p(x_C \mid x_B) \qquad (2.5)$$

moving fluidly here is (i.e.)

OR

$$p(x_A \mid x_B, x_C) = p(x_A \mid x_C) \qquad (2.6)$$

for all x_B : $p(x_B) > 0$

(2.5) - independence; with addition of conditioning on $|x_B$ formula

(2.6) - "probability of x_A given x_B and x_C is equal to probability of x_A given x_C" i.e. conditioning on x_B and x_C is the same as omitting x_B (conditioning variable) in prob.

- establish independence / conditional independence requires factorising the joint probability distribution

· PGM: representing a p.d. within g.m formalism involves making independence assumptions, which are embedded in the structure of the graph

· This graphical structure, other indep. relations can be derived, reflecting that certain factorisations of joint p.d. imply other factorisations

· Factorisations - read off via graph search algorithms

· Graphical structure encodes conditional independence:-

· chain rule of prob: PMF in a general factored form, given an ordering on $\{X_1, \ldots, X_6\}$:-

$$p(x_1, \ldots, x_6) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_1,x_2)\,p(x_4|x_1,x_2,x_3)\,p(x_5|x_1,x_2,x_3,x_4)\,p(x_6|x_1,x_2,x_3,x_4,x_5)$$

- under an arithmetic node ordering
- in general:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | x_1, \ldots, x_{n-1}) \qquad (2.7)$$

: compare (2.2) and (2.7) ⇒ conditioning variables dropped

missing variables in local c.p. functions ⇒ missing edges in underlying graph

It is this transfer of interpretation from missing variables → missing edges that underlies the probabilistic interpretation for missing edges in graph in terms of c.i.

· Formally: · Define an ordering I of nodes in a graph to be topological if for every node $i \in V$ the nodes in π_i appear before i in the ordering.

· E.g. $I = \{1,2,3,4,5,6\}$ (topological ordering for graph in 2.1)

· Let v_i denote set of nodes that appear earlier than i in the ordering I, excluding the parent nodes π_i

e.g. $v_5 = \{1,2,4\}$

· v_i necessarily contains ancestors of node i (excluding parents π_i) and may contain other nondescendant nodes also

· Given a topological ordering I for a graph G we associate to the graph the following set of c.i. statements:-

$$\{X_i \perp\!\!\!\perp X_{\nu_i} \mid X_{\pi_i}) \quad \text{for } i \in V \qquad (2.8)$$

· "Given the parents of a node (X_{π_i}), the node (X_i) is independent of all earlier nodes in the ordering (X_{ν_i})".

· The example encodes following indep. ass:-

$$\left(\begin{array}{l} X_1 \perp\!\!\!\perp \emptyset \mid \emptyset \\ X_2 \perp\!\!\!\perp \emptyset \mid X_1 \end{array} \right) \qquad \begin{array}{l} (2.9) \\ (2.10) \end{array}$$

$$X_3 \perp\!\!\!\perp X_2 \mid X_1 \qquad (2.11)$$

$$X_4 \perp\!\!\!\perp \{X_1, X_3\} \perp X_2 \qquad (2.12)$$

$$X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\} \perp X_3 \qquad (2.13)$$

$$X_6 \perp\!\!\!\perp \{X_1, X_3, X_4\} \perp \{X_2, X_5\} \qquad (2.14)$$

· Interpretation of missing edges as c.i. consistent with (2.2) - nuance later

· verification example of (2.12) by direct calculation from (2.3):-

mag. $\{X_1, \dots, X_4\}$ prob

$$p(x_1, x_2, x_3, x_4) = \sum_{x_5} \sum_{x_6} p(x_1, \dots, x_6) \quad \text{(marginalise)} \qquad (2.15)$$

$$\overset{\text{(subs.)}}{=} \sum_{x_5} \sum_{x_6} p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) p(x_5|x_3) p(x_6|x_2, x_5) \qquad (2.16)$$

$$\overset{\text{(sum.)}}{=} p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) \underbrace{\sum_{x_5} p(x_5|x_3)}_{=1} \underbrace{\sum_{x_6} p(x_6|x_2, x_5)}_{=1} \qquad (2.17)$$

$$= p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) \qquad (2.18)$$

mag. prob $\{X_1, X_2, X_3\}$:-

$$p(x_1, x_2, x_3) = \sum_{x_4} p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) \qquad (2.19)$$

$$= p(x_1) p(x_2|x_1) p(x_3|x_1) \sum_{x_4} p(x_4|x_2)$$

$$= p(x_1) p(x_2|x_1) p(x_3|x_1) \qquad (2.20)$$

Hence: $p(x_4 | x_1, x_2, x_3) = p(x_4 | x_2)$ (recall alt c.i. definition!) (2.21)

(Q) Above shows we can interpret missing edges in graph in terms of conditional independencies

(*) (Q): are there other conditional independence statements that are true of such joint probability distri's ; and do these have graphical interp.

- There are other conditional independencies e.g. $X_1 \perp\!\!\!\perp X_6 \mid \{X_2, X_3\}$, not in list (2.14), but implied by the list.
- verifiable by algebra (tedious)

(Q): we want to write down ALL conditional independencies implied by basic set.

- Develop criterion for doing so without factorising joint into all possible ~~tripl~~ triples of var. subsets

(Q) That is, a general graph search algorithm to ~~so~~ find all implied independencies as well as explicit ones from graph structure.

2.12. Conditional independence and Bayes ball

- Bayes ball - reachability under a def. of separation.
- We can not only derive C.I. assertions from (2.2); but other way around.

Relevance:
- To each graph we associate family of joint prob. distri.
- Family arises due to consideration of range over different choices of numerical values of local c.p.s $p(x_i \mid x_{\pi_i})$. list of
- View c.i. statements gen by Bayes ball as constraints on ~~F~~ joint p.d.
- Those joint p.d.s that meet → in the family, if not out.

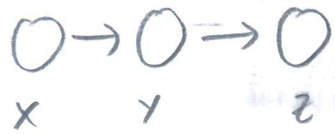
(Q): Relationship between characterisation of a family of p.d.s in terms of conditional independences ; and numerical characte. in terms of local cond. prob. (5.2.1.3.)

---

3 canonical graphs

missing edges in PGM → c.i.

le



(2.4)

- Missing edge: X–Z
- Hence C.I:–

① $X \perp\!\!\!\perp Z \mid Y$

② No other conditional independencies associated with this graph

- Justifying ①: (DAG structure ⇒ C.I.)

From DAG: $p(x,y,z) = p(x)p(y|x)p(z|y)$ (2.23)
(2.2)

$$\Rightarrow p(z|x,y) = \frac{p(x,y,z)}{p(x,y)} \quad (2.24)$$

$$= \frac{p(x)p(y|x)p(z|y)}{p(x)p(y|x)} \quad (2.25)$$

$$= p(z|y) \quad (2.26)$$

- Establishing c.i.
- Justifying ②:– (SN)
- "There are no further conditional independencies associated with this graph"
  - Does NOT mean that no further conditional independencies can arise in any of the distributions in the family associated with this graph i.e. distributions with factorised form (2.23) (?)
  - There exist some distributions which exhibit conditional independencies
  - e.g. free to choose any local c.p. $p(y|x)$ → choose distn in which all prob. y same no matter value of x. then for this $p(y|x)$, $X \perp\!\!\!\perp Y$.

(SN): Fig(2.4) ⇏ X and Y are necessarily dependent (not independent).

(SN): Edges do not necessarily imply dependence

BUT (SN): Missing edges do imply independence

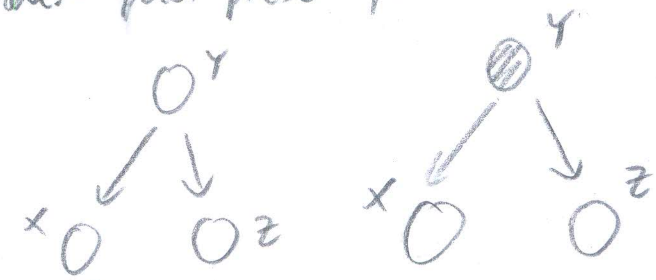
- universally vs existentially qualified statements with respect to family of distributions associated with a graph.
- (SN): Asserted conditional independencies always hold for these distributions
  - implied/non-asserted conditional independencies sometimes fail to hold for distributions associated with a particular graph, but sometimes do hold.

(3): · An algorithm based on conditional independencies will be correct for all distr's associated with a graph
· An algorithm based on <u>absence</u> of c.i.s. will sometimes be correct, sometimes not.

<u>Mental model</u> for cascade:- past-present-future

<u>common parent</u>

-<u>Missing edge</u>: X-Z



(2.5

(1) <u>$X \perp\!\!\!\perp Z \mid Y$</u>

· Justification (1):- (DAG → C.I.)

$$p(x,y,z) = p(y)p(x|y)p(z|y) \quad (2.28)$$

$$p(x,z|y) = \frac{p(y)p(x|y)p(z|y)}{p(y)} \quad (2.29)$$

$$\Rightarrow p(x,z|y) = p(x|y)p(z|y) \Rightarrow X \perp\!\!\!\perp Z \mid Y \quad (2.30)$$

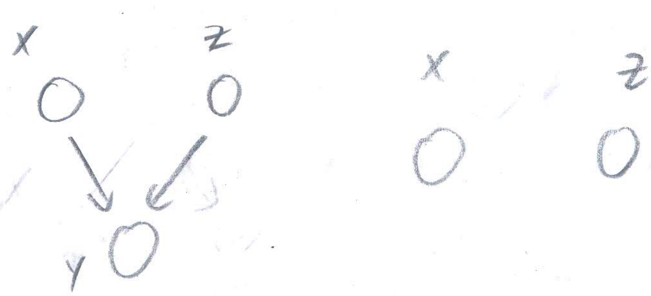
<u>Mental model:</u> 'Hidden variable'

- X - shoe size ; Z - 'gray hair' ; Y - age
- X, Z → in pop., strong dep.
- But with Y - age, we might be willing to assert $X \perp Z \mid Y$.
- Hidden 'Y' explains all of <u>observed X-Z dep.</u>

(2) "No other c.is associated" with this graph"

- No assertions of <u>dependence</u>, in particular we do not necess. assume X and Z dependent because they depend on Y.
- But we assert that at least some distr's in which such dependencies
<u>& found.</u>

<u>V-structure</u>



$X \perp\!\!\!\perp Z$

- Missing edge X-Y

(Statement about subgraph)

10 - PGMS readings

Koller + Friedman (2009)

- (OT): use the intuitive examples here to supplement, reinforce understanding; they hold the key to applying the reasoning
- (W): The purposes of this reading → clarifying BN semantics, I-maps, equivalence, soundness, completeness

- (W): recall 36-705, IID as cornerstone of statistical analysis
- (W): 10-708 → complementary through consideration of dependence

---

- Reason for parametrisation through cpd rather than full joint → modularity
- Adding node will not require new joint p.d.; modular in cpd representation.
- local vs global

3.2.2.2. Bayesian Network semantics

Definition 3.1

- A Bayesian Network structure G is a DAG whose nodes represent r.v.s. $X_1, \ldots, X_n$.
- Let $Pa^G_{X_i}$ represent parents of X_i in G
- Let $NonDescendants_{X_i}$ denote variables that are not descendants of X_i
- G encodes following set of conditional independence assumptions, called local independencies.
- These are denoted $I_\ell(G)$:-

For each variable X_i : $(X_i \perp\!\!\!\perp NonDescendants_{X_i} \mid Pa^G_{X_i})$

"The local independencies state that each node X_i is conditionally independent of its nondescendants given parents"

3.2.3. Graphs and Distri

(G):- 1.) Bayesian Network graph with conditional independence assertions

2.) Bayesian Network graph with conditional probability distributions

These definitions are equivalent

(07):

(*) Distribution P satisfies local independencies associated with a graph G $\Leftrightarrow$ P is representable as a set of CPDs associated with graph G

· 3.2.3.1. - I maps

- define set of independencies associated with distri P

Definitions 3.2 - independencies in P

- Let P be a distri over X

(*) we define $I(P)$ to be the set of independence assertions of the form $(X \perp Y \mid Z)$ that hold in P.

(*) - we can rewrite statement that "P satisfies the local independencies associated with G" as $I_\ell(G) \subseteq I(P)$

(*) - semantically, "G is an I-map (independency map) for P"

Definition 3.3 - I-map

- Let K be any graph object associated with a set of independencies $I(K)$
- We say "K is an I-map for a set of independencies I if $I(K) \subseteq I$"

(*) - Semantically

"G is an I-map for P if G is an I-map for $I(P)$"

(Q): do you understand these nuances?

- (*) Direction of subset inclusion $\Rightarrow$ for G to be an I-map of P, it is necessary that G does not mislead us regarding independencies in P.

- Stare at example to help with this abstract definition

- Note that $I_\ell(G) \subseteq I(P)$ implies that $I(P)$ may contain independence assertions not present in $I_\ell(G)$!

- These definitions are finnicky $\rightarrow$ do some scribblings

Intuition from I-map to factorisation

write statements mathematically

- BN structure G encodes a set of c.i assumptions
- Every distri for which G is an I-map must satisfy the assumptions
- This is key to understanding factorised representation
- Consider any distri P for which student BN $G_{student}$ is an I-map.

- Decompose joint distribution via chain rule:-

$$P(I,D,G,L,S) = P(I)P(D|I)P(G|I,D)P(L|I,D,G)P(S|I,D,G,L) \quad (3.15)$$

- No assumptions, but not helpful
- Apply c.i. assumptions, induced by B.N.
- Assume $G_{student}$ is an I-map for our distri. P

$(3.14) \rightarrow (I \perp D) \in I(P) \Rightarrow P(D|I) = P(D)$

$(3.10) \rightarrow (L \perp I, D | G) \in I(P) \Rightarrow P(L|I,D,G) = P(L|G)$

$(3.11) \rightarrow (S \perp D,G,L | I) \in I(P) \Rightarrow P(S|D,G,L,I) = P(S|I)$

- Hence $P(I,D,G,L,S) = P(I)P(D)P(G|I,D)P(L|G)P(S|I) \quad (3.16)$

- Any entry in joint distri can be computed as a product of factors, one for each var.
- Each factor $\longrightarrow$ a conditional prob. of variable given network parents
- This factorisation applies to any distribution P for which $G_{student}$ is an I-map

Formally:

Def 3.4 (factorisation)

- Let G be a BN graph over $X_1, \dots, X_n$
- We say a distri P over same space factorises according to G if P can be expres:-

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | Pa^G_{X_i})$$

- chain rule for BN
- individual factors $P(X_i | Pa^G_{X_i})$ are CPDs (conditional prob distri.)

03.5. Bayesian Network

- A Bayesian Network is a pair $B=(G,P)$ where P factorises over G, and where P is specified as a set of CPDs associated with G's nodes
- The distri P is often annotated P_B.
- Phenomenon holds for Gstudent more generally.

Theorem (3.1)

- Let G be a BN structure over a set of r.v.s. X, and let P be a joint distribution over the same space.

(*) If G is an I-map for P, then P factorises according to G.

Proof: - see Koller & Friedman (may use for exercises)

comments: - C.I. assumptions implied by BN structure G allow factorisation of a distri P for which G is an I-map into small CPDs.

- (G): 1.3.1. shows direction of fundamental connection between the c.independencies encoded by the BN structure and the factorisation of distri into local probabilistic models / CPDs.

- C.I. $\Rightarrow$ factorisation

- Theorem 3.2 (converse: factorisation according to $G \Rightarrow$ C.I.)

- Let G be a BN structure over a set of r.v.s. X and let P be a joint distri. over same space.

(*) If P factorises according to G, then G is an I-map for P.

- 3.3 Independencies in graphs

- Graph structure $G \rightarrow$ encodes a set of c.i. assumptions $I_\ell(G)$
- Knowing only that a distribution factorises over G; we can conclude that distri satisfies $I_\ell(G)$. (i.e. $I_\ell(G) \subseteq I(P)$)
- (Q): Recall Jordan : - Are there any other independencies that we can 'read off' directly from G.
  - Are there other independencies that hold for every distribution P that factorises over G? (in distri family)

(A): A question about implied independencies i.e. $I(P) - I_\ell(G)$ (?)

D-separation

- When can we guarantee that an independence $(X \perp Y | Z)$ holds in a distri associated with a BN structure G.
- In interests of concision, I'm going to leave comments
- The next section is on the canonical local structures / independencies; in Jordan, cascade, common parent, v-structure.
- Koller (2009) adds additional info / intuition and categorises as:-

1. Direct connection $X \to Y$
2. Indirect connection (3-node networks)
   - i) Indirect causal
   - ii) Indirect evidential
   - iii) common cause
   - iv) common effect

   } each contains intuition

- Key is to note that the key event is whether a node is observed or unobserved (around which C.I statements (dependencies) / uncoupling is generated)
- Koller (2009): (N): when probabilistic influence can flow from X to Y via Z; then we have an active trail / blocking
- 10708 L2 notes: Better summary linking Bayes Ball to Koller's idea of active trails.

(N): essentially: { Bayes Ball - reachability - ball - blocked (Jordan)
- probabilistic influence - active - trails - blocked (Koller)

→ Both encode intuitions about what leads to d-separation

General case: · Longer trail $X_1 \rightleftarrows \dots \rightleftarrows X_n$
· For influence to flow from X_1 to X_n; it needs to flow through every node on trail.
· X_1 can influence X_n if every two edge trail. $X_{i-1} \rightleftarrows X_i \rightleftarrows X_{i+1}$ along trail allows influence to flow.

Formally:

D.3.6.

Let G be a BN structure; and $X_1 \rightleftarrows X_2 \rightleftarrows \dots \rightleftarrows X_n$ a trail in G.
Let Z be a subset of observed r.v.s.
The trail $X_1 \rightleftarrows X_2 \rightleftarrows \dots \rightleftarrows X_n$ is active given Z if:-

- whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ then X_i or one of its descendants are in $\underline{Z}$.
- No other node along the trail is in $\underline{Z}$

- If X_1 or X_n are in $\underline{Z}$, the trail is not active.

- Need to account for graphs with more than one trail between nodes
- d-separation (*): d-separation
  - i) Bayes ball algorithm
  - ii) separation on moralised ancest. graph
  - iii) Active trails
  - (iv). Multiple algos for same intuition

D.3.7 (d-separation)

- Let $\underline{X}, \underline{Y}, \underline{Z}$ be three sets of nodes in G.
- We say $\underline{X}$ and $\underline{Y}$ are d-separated given $\underline{Z}$ i.e. $dsep_G(\underline{X} : \underline{Y} | \underline{Z})$, if there is no active trail between any nodes $X \in \underline{X}$ and $Y \in \underline{Y}$ given $\underline{Z}$.
- $I(G)$ is the set of independencies that correspond to d-separation

$$I(G) = \{(\underline{X} \perp \underline{Y} | \underline{Z}) : d\text{-}sep_G(\underline{X} : \underline{Y} | \underline{Z})\}$$

This set is called the set of global Markov independencies
The independencies in $I(G)$ are precisely those that are guaranteed to hold for every distri over G.

some formalisms (mathematical) regarding d-sep. (i.e. proof style)
desirable properties of d-separatn

3. (Soundness of d-sep)

If a distri P factorises according to G, then $I(G) \subseteq I(P)$

mark: - Any independence reported by d-separation is satisfied by the underlying distri
- If two nodes X and Y are d-separated given some $\underline{Z}$, we are guaranteed they are c.i. given $\underline{Z}$.

pleteness - d-separation detects all possible independencies
- If two variables X and Y are independent given $\underline{Z}$, they are d-separated
- defined in this form $\longrightarrow$ no specification of distribution in which X and Y are independent
- e. actually a statement which involve a set of distri / distri family?

D 3.8. (faithful)

- A distri P is faithful to G if, whenever $(X \perp Y \mid Z) \in I(P)$; then $dsep_G(X; Y \mid Z)$.
- Any independence in P is reflected in the d-sep. properties of graph (?)

(Koller) (Xing)

- A candidate formalisation of completeness is in terms of converse/contrapositive

(*) for any distri P that factorises over G, we have that P is faithful to G

(*) i.e. if X and Y are not d-separated given Z in G, then X and Y are dependent in all distribut. P that factorise over G. (1/2)

(?) :- converse to soundness (Koller)
- If true, two together (which?) imply that for any P that factorises over G, we have $I(P) = I(G)$.

- Highly desirable property is false (?) which (?)

(*) :- Even if a distri factorises over G, it can still contain additional independencies
(*) not reflected in structure

- completeness property does not hold for this candidate definition of completeness
- settle for a weaker definition:-
  - If $(X \perp Y \mid Z)$ in all distrib P that factorise over G, then $dsep_G(X : Y \mid Z)$.
  - contrapositive :- If X and Y are not d-separated given Z in G, the X and Y are dependent in some distribution that factorises over G.

- Formally construct theorem:

1.3.4

- Let G be a BN structure.
- If X and Y are not d-separated given Z in G, then X and Y are dependent given Z in some distribution that factorises over G.

Proof:- see Koller (2009).

- Remark: completeness result tells us that our definition of $I(G)$ is maximal. For any independence assertion not a consequence of d-separation in G, we can find a counterexample distri P that factorises over G.

T3.5 (measure theoretic qual.)

- For almost all distributions P that factorise over G, that is for all distributions except for a set of measure zero in the space of CPD parametrisations, we have $I(P) = I(G)$.

· Stronger than 1.3.4

(*) (*) - Results state that for almost all parametrisations P of the graph G
(that is for almost all possible choices of CPDs for the variables), the d-separation test precisely characterise the independences that hold for P.

3.3.4 - I-equivalence

· I(G) specifies a set of c.i. assertions associated with a graph.

· can abstract away details of graph structure and view as specification of indep. properties

(*) - very different BN structures can be equivalent in that they encode precisely the same set of c.i assertions
(e.g. local structures/canonical maps).

D.3.9. (I-equivalence)

- two graph structures K_1 and K_2 over X are I-equivalent if $I(K_1) = I(K_2)$

- The set of all graphs over X is partitioned into a set of mutually exclusive and exhaustive I-equivalence classes.

- which ~~is the~~ are the set of equivalence classes induced by the I-equivalence relation

(*) I-equivalence of 2 graphs ⟹ any distr P that can be factorised over one of these graphs can be factorised over the other.

(*) - No intrinsic property of P that would allow us to associate it with one graph rather than an equivalent one.

(*) - important implications on our ability to determine dir. of influence

10-708 22/05/2020

L2 - Bayes ball / d-sep. (suppl.)

(*) 2 main perspectives; both encoding same set of rules on d-separation
- Jordan (2003): Bayes-ball algorithm / blocking
- Koller (2009): probabilistic influence / active trails

(*) d-separation has to account for naive graph sep. and fails with V-structures.

Jordan (2003): Ch 2

(*) Decide whether a given condit. independence statement $X_A \perp\!\!\!\perp X_B | X_C$ is true for a directed graph G.

(*) Formally; this means that the statement holds for every distribution that factors according to G (later).

(*) Use 3 canonical graphs (local structures)

(*) Reachability algorithm:-
- shade node you are conditioning on i.e. X_C
- Place a ball at one of r.v.s i.e. X_A (source)
- send it to the other r.v. i.e. X_B (destination)

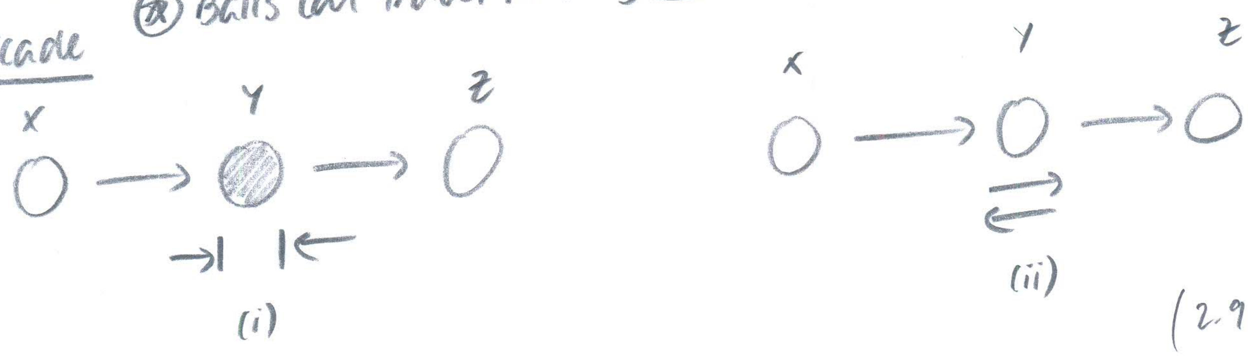
(*): Does a ball reach the destination X_B from source X_A?

YES $\longrightarrow$ $X_A \perp\!\!\!\perp X_B | X_C$ is not true

NO $\longrightarrow$ $X_A \perp\!\!\!\perp X_B | X_C$ is true.

(*) Bayes ball specifies a set of rules on ball movement
(*) Balls can travel in any dir along directed edges

- cascade



(2.9)

(i) Assert $X \perp\!\!\!\perp Z | Y$

(ii) Do not assert $X \perp\!\!\!\perp Z | Y$

• For this structure; ball is blocked when conditioning

common parent

(2.10)

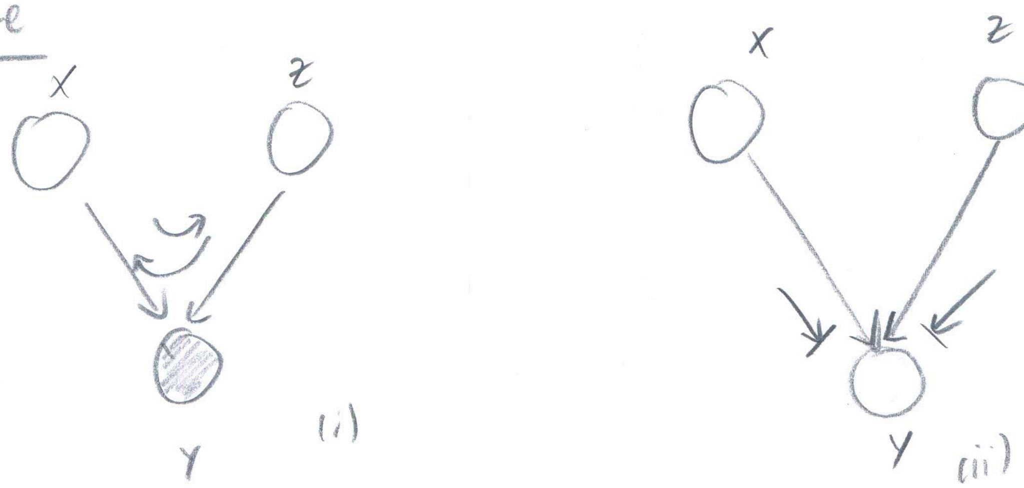
i) Assert $X \perp\!\!\!\perp Z \mid Y$

ii) Do not assert $X \perp\!\!\!\perp Z \mid Y$

· For this structure, ball is <u>blocked</u> when conditioning.

(*) Note for cascade and common parents; conditioning on a node has the effect of blocking balls. This does not extend to V-structures; as Jordan points out.

V-structure



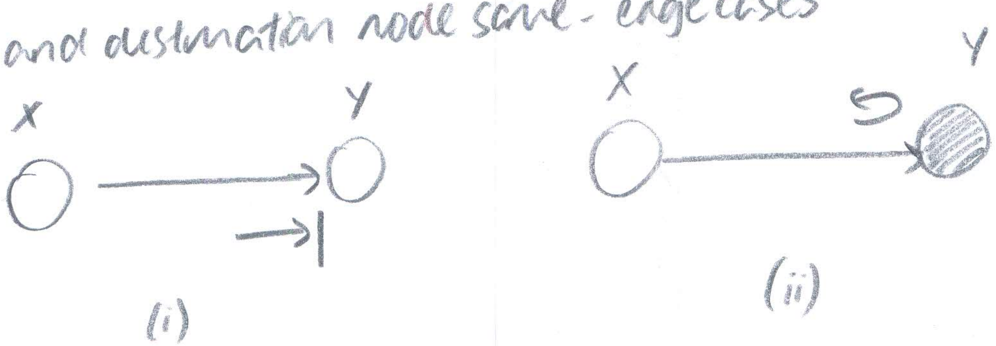
(2.11)

i) Do not assert $X \perp\!\!\!\perp Z \mid Y$

ii) Assert $X \perp\!\!\!\perp Z \mid Y$

- For this structure, ball <u>passes through</u> when conditioning.

(*) source and destination node same - edge cases



(2.12)

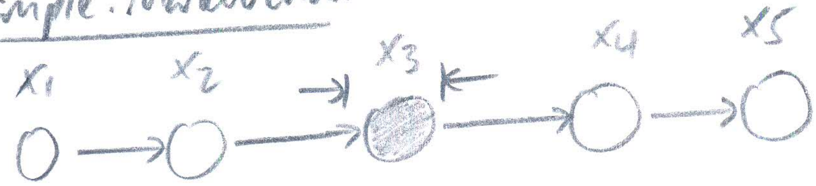
· ~~on~~ For this structure; ball <u>passes through</u> when conditioning

(*) Source and destination node - edge cases

(2.13)

- For this structure; ball is <u>blocked</u> ~~when~~ <u>conditioning</u>

(*) <u>Example: Markov chain</u>



- without d-sep:-

$$p(x_1, x_2, \ldots, x_6) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2)\,p(x_4|x_3)\,p(x_5|x_4)$$

- arbitrarily (for ordering I=1,2,3,4,5,6)

$$p(x_1, \ldots, x_6) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2,x_1)\,p(x_4|x_3,x_2,x_1)\,p(x_5|x_4,x_3,x_2,x_1)$$

- Hence; <u>without d-sep</u>:-

$$X_5 \perp\!\!\!\perp \{X_3, X_2, X_1\} \mid X_4$$

$$X_4 \perp\!\!\!\perp \{X_2, X_1\} \mid X_3$$

$$X_3 \perp\!\!\!\perp X_1 \mid X_2$$

$$X_2 \perp\!\!\!\perp \emptyset \mid X_1$$

$$X_1 \perp\!\!\!\perp \emptyset \mid \emptyset$$

(*) with <u>d-sep</u>:

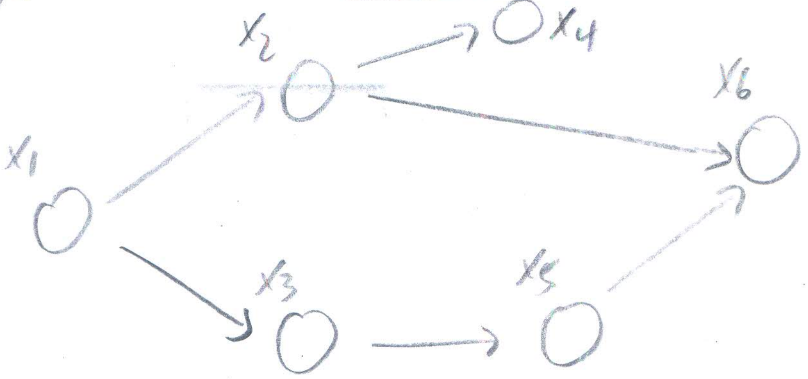
- observe, as an example; that X_3 blocks subsets $\{X_1, X_2\}$ and $\{X_4, X_5\}$.
- yielding some additional condit. indep:-
  (not exhaustive list) (using cascade + d-sep)

$$X_1 \perp\!\!\!\perp X_5 \mid X_4 \qquad X_1 \perp\!\!\!\perp X_5 \mid X_2 \qquad X_1 \perp\!\!\!\perp X_5 \mid \{X_2, X_4\} \qquad \text{etc.}$$

- Bayes ball shortcuts algebraic manipulation

(*) In general, in this cascade, with time as a mental model, ~~we can~~ any subset of 'future' nodes is conditionally indep. of any subs 'past' nodes given the subset that <u>blocks</u> them.

(*) Bayes ball - another composite example



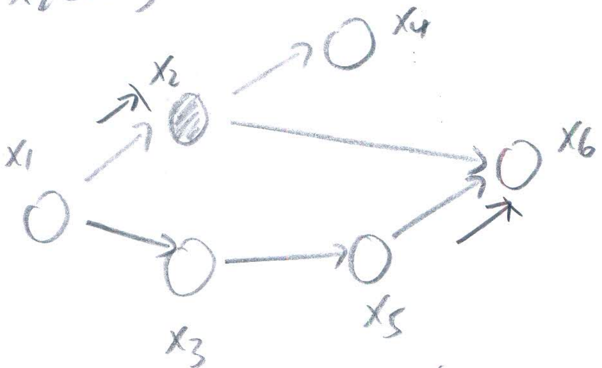
- other than existing (C.I) statements in (2.9)-(2.14), we have further C.I. statements

- i) $X_4 \perp\!\!\!\perp \{X_1, X_3\} \mid X_2$

  ii) $X_1 \perp\!\!\!\perp X_6 \mid \{X_2, X_3\}$

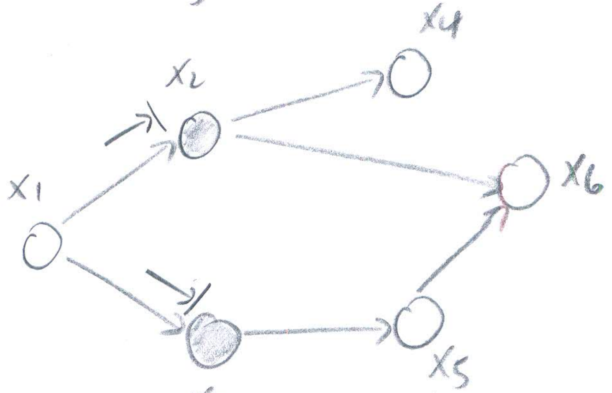
  iii) $X_2 \perp\!\!\!\perp X_3 \mid \{X_1, X_6\}$ not true

i)



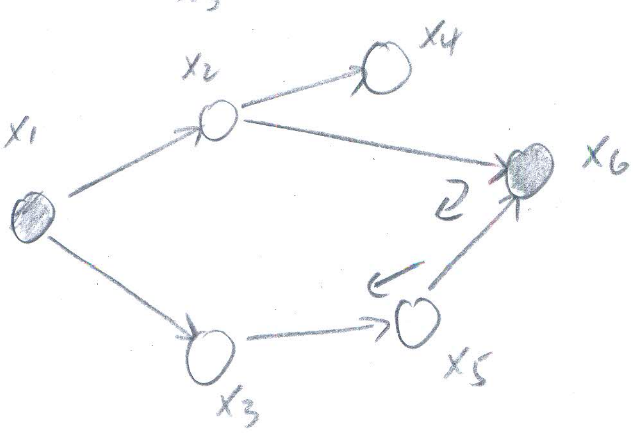
via 2.9 (i)
2.12 (i)

ii)



via 2.9 (i)

iii)



2.9(i)
via 2.11 (i)

- obs. of X_6 → explaining away dep between X_2, X_5 (v-struc.)

- X_5 and X_3 dep. (?)
- X_2, X_3 dep.