Graphical Models and Markov Blankets

Stephan Stahlschmidt

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin

Why Graphical Models?

Illustration

Highlighting the probabilistic dependence structure:

- *Nodes* represent variables
- *Edges* state dependence between variables
- *Missing edges* encode independencies

Statistical Inference

Entering of evidence allows for inference

Outline

1. Motivation ✓
2. Basic Concepts
3. Inference
4. Correctness and Completeness
5. Learning from Data
6. Application

Basic Graph Theory

Directed Graph	Edges indicate a direction via an arrow
(Directed) Path	Sequences of subsequent edges (in the same direction)
Directed Cycle	Directed path with identical start and endpoint
DAG	Directed Acyclic Graph
Parents of a node X	Nodes pointing to X via a directed edge
Ancestors	Parents and all subsequent parents
Children of a node X	Nodes pointed at from X via a directed edge
Descendants	Children and all subsequent children

Basic Probability Calculus

Multiplication Rule

$$P(X, Y) = P(X|Y)\,P(Y) = P(Y|X)\,P(X)$$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{p} P(X_i|X_1, \ldots, X_{i-1})$$

Independence $X \perp\!\!\!\perp Y$

$$P(X|Y) = P(X)$$
$$P(X, Y) = P(X)\,P(Y)$$

Conditional Independence $X \perp\!\!\!\perp Y|Z$

$$P(X|Y, Z) = P(X|Z)$$
$$P(X, Y, Z) = P(X|Z)\,P(Y|Z)\,P(Z)$$

Decomposition

Assume pdf $\mathrm{P}(A, B, C, D, E)$ with

- $A \perp\!\!\!\perp B | E$
- $B \perp\!\!\!\perp C | D$
- $C \perp\!\!\!\perp D | E$
- $D \perp\!\!\!\perp E$

Decomposition

1. Factorise pdf via the Multiplication Rule:

$$\begin{aligned}\mathrm{P}(A,B,C,D,E) = &\ \mathrm{P}(A|B,C,D,E)\,\mathrm{P}(B|C,D,E)\,\mathrm{P}(C|D,E)\\ &\ \mathrm{P}(D|E)\,\mathrm{P}(E)\end{aligned}$$

2. Reduce complexity due to independence in factors:

$$\mathrm{P}(A,B,C,D,E) = \mathrm{P}(A|C,D,E)\,\mathrm{P}(B|D,E)\,\mathrm{P}(C|E)\,\mathrm{P}(D)\,\mathrm{P}(E)$$

Decomposition and DAGs

3. Describe the decomposed pdf via a DAG:
   (a) Draw a node for every factor
   (b) Add edges according to the conditional statements in the factors

Decomposition and DAGs

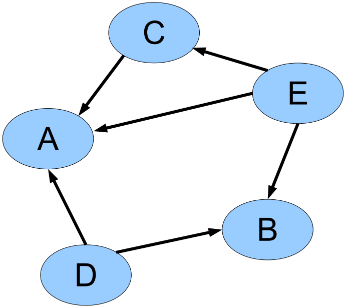


Figure 1: DAG for the decomposed pdf $\mathrm{P}(A, B, C, D, E) = \mathrm{P}(A|C, D, E)\,\mathrm{P}(B|D, E)\,\mathrm{P}(C|E)\,\mathrm{P}(D)\,\mathrm{P}(E)$

Bayesian Networks

A Bayesian Network consists of a

- DAG (*structure*)
- Corresponding (conditional) pdf for every node (*parameter*)

Generalisation of Bayes rule:

$$P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$

Statistical Inference

- Initialised by introducing evidence in a specific node
- Information is sent via the edges to the parents and children
- Causes probability altering throughout the network until the information flow is blocked (d-seperation)

Reasoning with Graphical Models

Three kind of reasoning arise:

1. **Deductive Reasoning:** Information follows the edges' direction
2. **Abductive Reasoning:** Information flow contradicts the edges' direction
3. **Inter-Causal Reasoning:** Evidence supporting a certain hypothesis will increase that probability and decrease the probability of the alternative hypothesis ("explaining away")

Information flow in Graphical Models

- **Serial Connection:** $A \longrightarrow C \longrightarrow B$
  Information flows freely, as long as no evidence is entered in the middle node
- **Diverging Connection:** $A \longleftarrow C \longrightarrow B$
  Information is transmitted unless evidence is entered in the middle node
- **Converging Connection:** $A \longrightarrow C \longleftarrow B$
  Information flow is blocked unless evidence is entered for the middle node or its descendants (Collider)

d-separation

Feasibility of information flow between two variables in a DAG

Every path between nodes A and B is blocked by a set $\mathcal{C}$, which is on this path, if

- evidence is entered for node $C \in \mathcal{C}$ and the arrows do not meet head-to-head on C or
- no evidence is entered for $C \in \mathcal{C}$, but the arrows meet head-to-head on C (Converging Connection)

In these cases A is d-separated from B: $A \perp B$

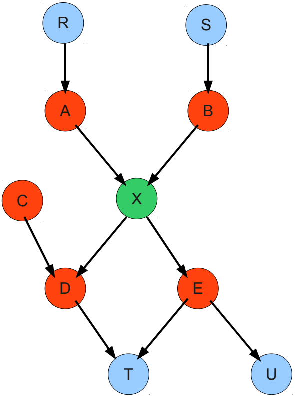
Markov Blanket

The Markov Blanket MB(X) shields the variable X against any information.

It consists of

- All parents of X
- All children of X
- All further parents of X's children

If the state of all variables in the MB(X) is known, no other variable $Y \notin$ MB(X) can influence X.



Correctness and Completeness

Varying the variable order in the multiplication rule leads to different factors in the subsequent decomposition

Probability distribution may be too complex to be mirrored perfectly by a DAG

Correctness and Completeness

Assume pdf P over a set of variables V, which includes the subsets A, B and S. $\mathcal{G}$ represents a DAG with the corresponding X_A, X_B and X_S.

Correctness of $\mathcal{G}$ with respect to P (I-map):

$$X_A \perp X_B | X_S \Longrightarrow A \perp\!\!\!\perp B | S$$

Completeness of $\mathcal{G}$ with respect to P (D-map):

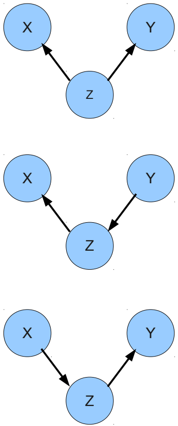
$$X_A \perp X_B | X_S \Longleftarrow A \perp\!\!\!\perp B | S$$

A correct and complete DAG $\mathcal{G}$ is a **perfect map** of a pdf P.

Variable Order

Assume $\mathrm{P}(X, Y, Z)$ with $X \perp\!\!\!\perp Y | Z$

$$
\begin{aligned}
\mathrm{P}(X, Y, Z) &= \mathrm{P}(X|Y, Z)\,\mathrm{P}(Y|Z)\,\mathrm{P}(Z) \\
&= \mathrm{P}(X|Z)\,\mathrm{P}(Y|Z)\,\mathrm{P}(Z) \\
\mathrm{P}(X, Z, Y) &= \mathrm{P}(X|Z, Y)\,\mathrm{P}(Z|Y)\,\mathrm{P}(Y) \\
&= \mathrm{P}(X|Z)\,\mathrm{P}(Z|Y)\,\mathrm{P}(Y) \\
\mathrm{P}(Y, Z, X) &= \mathrm{P}(Y|Z, X)\,\mathrm{P}(Z, X)\,\mathrm{P}(X) \\
&= \mathrm{P}(Y|Z)\,\mathrm{P}(Z|X)\,\mathrm{P}(X)
\end{aligned}
$$

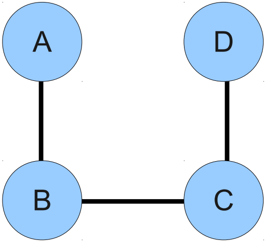


Graphs are **Markov equivalent**.

Complex pdf

Assume $\mathrm{P}(A, B, C, D)$ with

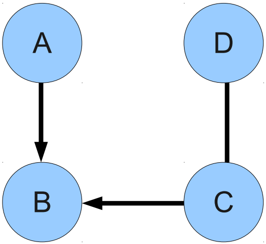
- $A \perp\!\!\!\perp D$
- $A \perp\!\!\!\perp C | B, D$
- $D \perp\!\!\!\perp B | A, C$



Complex pdf

Assume $\mathrm{P}(A, B, C, D)$ with

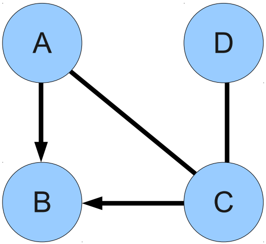
- $A \perp\!\!\!\perp D$
- $A \perp\!\!\!\perp C | B, D$
- $D \perp\!\!\!\perp B | A, C$



Complex pdf

Assume $\mathrm{P}(A, B, C, D)$ with

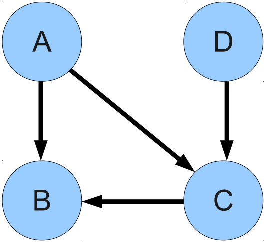
- $A \perp\!\!\!\perp D$
- $A \perp\!\!\!\perp C | B, D$
- $D \perp\!\!\!\perp B | A, C$



Complex pdf

Assume $P(A, B, C, D)$ with

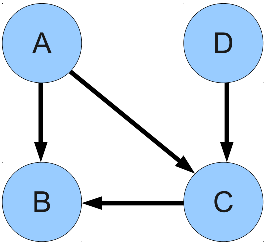
- $A \perp\!\!\!\perp D$
- $A \perp\!\!\!\perp C | B, D$
- $D \perp\!\!\!\perp B | A, C$



Complex pdf

Assume $\mathrm{P}(A, B, C, D)$ with

- $A \perp\!\!\!\perp D$
- $A \perp\!\!\!\perp C | B, D$ $\rightarrow$ not included in incomplete, but correct DAG
- $D \perp\!\!\!\perp B | A, C$



Markovian Parents

Markovian parents PA_i denote a minimal set of predecesors of X_i, that make X_i independent of all other predecessors.

Assuming an ordered set of variables:

$$\begin{aligned}
\mathrm{P}(X_1, \ldots, X_p) &= \prod_{i=1}^{p} \mathrm{P}(X_i | X_1, \ldots, X_{i-1}) \\
&= \prod_{i=1}^{p} \mathrm{P}(X_i | PA_i)
\end{aligned}$$

Markov Compatibility

Graph $\mathcal{G}$ and pdf P are Markov compatible, if

- P can be factorised according to Markovian parents
- Factorisation can be drawn as a DAG

Markov compability is a necessary and sufficient condition for a DAG $\mathcal{G}$ to become a Bayesian Network of a pdf P.

Learning a Bayesian Network

Need to learn:

- **Structure**: np-complete
  10 variables can be assembled into $\sim 10^{18}$ different DAGs
- **Parameters**: easy
  e.g. maximum likelihood estimation for (conditional) probabilities in the nodes

Structure Learning

Two approaches:

1. **Score-based methods**: Evaluate a score for DAGs in a narrowed search space
2. **Constraint-based methods**: Local (conditional) independence tests
   Assumption: Correct independence tests and faithfulness

Faithfulness

Bayesian Network is faithful iff every d-connection in $\mathcal{G}$ is mirrored by a conditional dependence in P:

$$X_A \not\perp X_B | X_S \Longrightarrow A \not\perp\!\!\!\perp B | S$$

Graphical Models and Markov Blankets

Stephan Stahlschmidt

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin

References

D. Edwards
Introduction to Graphical Modelling
Heidelberg: Springer Verlag, 2000

U. Kjærulff and A. Madsen
Bayesian Networks and Influence Diagrams
Heidelberg: Springer Verlag, 2008

N. Meinshausen and P. Bühlmann
Stability Selestion
arXiv: 0809.2932v1

References

- R. Neapolitan
  Learning Bayesian Networks
  New York: Pearson Prentice-Hall, 2003
- J. Pearl
  Causality
  Cambridge: Cambridge University Press, 2000
- P. Spirtes, C. Glymour and R. Scheines
  Causation, Prediction, and Search
  Cambridge: MIT Press, 2001