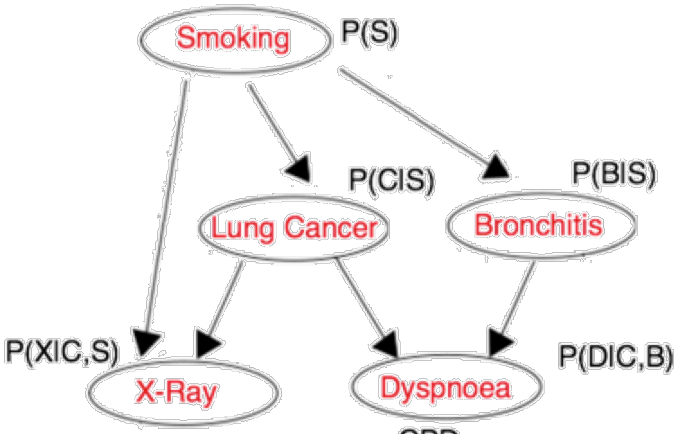
ETH zürich

Tutorial: Learning Bayesian Networks with R January 2018

Polina Suter

polina.minkina@bsse.ethz.ch

Bayesian Networks



CPD:

C	B	D=1	D=0
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

- G – Directed Acyclic Graph (DAG)
- Nodes – random variables
- Edges – conditional dependencies
- CPD (conditional probability distribution) of node X:

  P(X|parents(X))

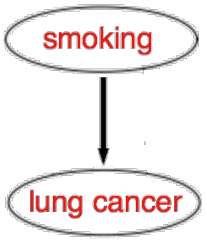
Compact representation of joint distribution in a product form:

$$P(S,C,B,X,D)=P(S)P(C|S)P(B|S)P(X|C,S)P(D|C,B)$$

$1+2+2+4+4=13$ parameters instead of $2^5=32$

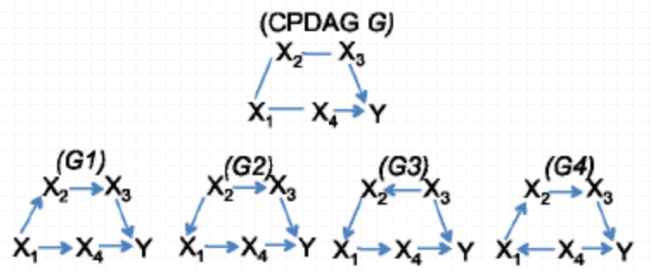
Learning BNs from observations/interventions

- The gold standard for inferring causal links is intervention experiment however it is not always possible to perform one

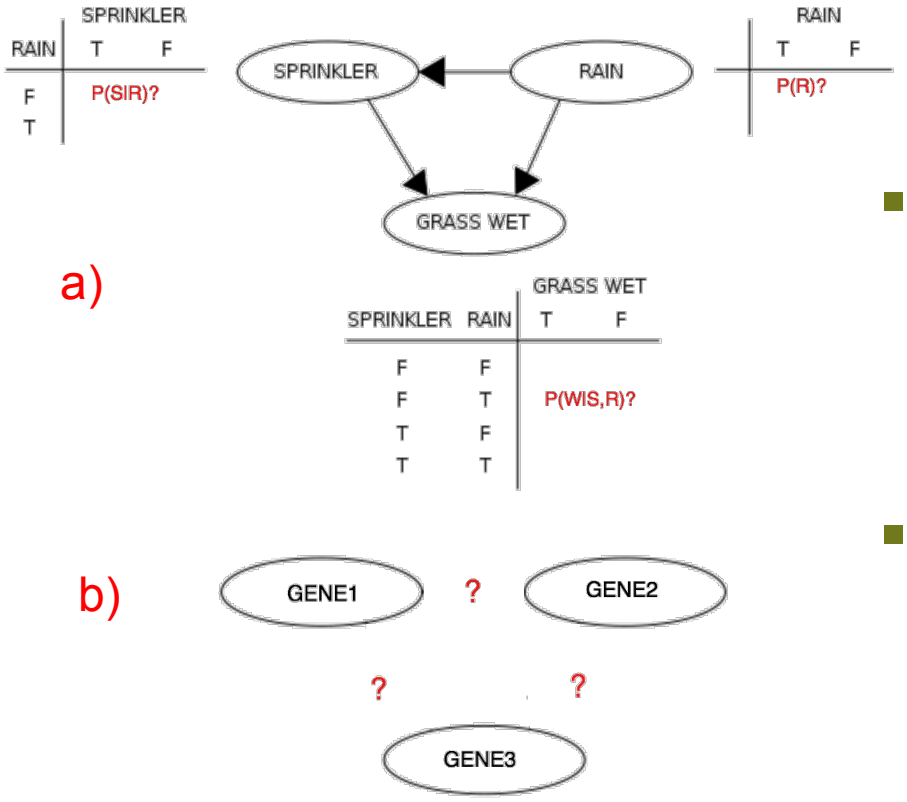


- Inferring BNs from observational data has some limitations:

1. For causal interpretation we assume there are no hidden confounders
2. Only equivalence class (CPDAG) can be inferred from observational data



Learning BNs



$$\hat{G} = \arg\max_{G} Score(G)$$

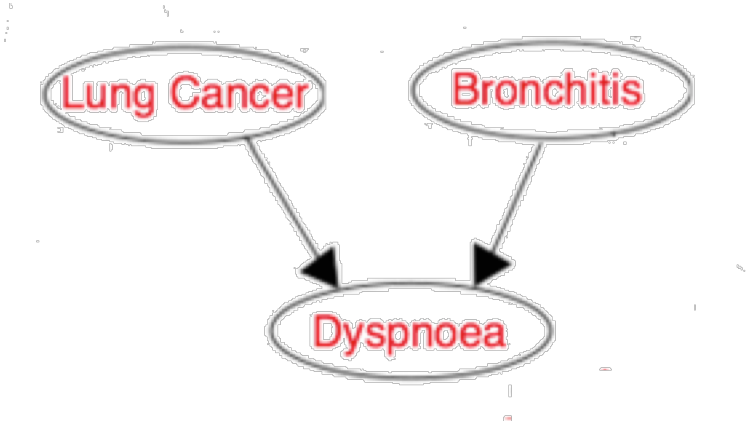
NP-hard

- Learn parameters Θ when structure G is known: ML, MAP
- Learn graph G and parameters Θ

p	number of DAGs with p nodes
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	1213442454842881
10	4175098976430598143
11	31603459396418917607425
12	521939651343829405020504063
13	18676600744432035186664816926721
14	1439428141044398334941790719839535103
15	237725265553410354992180218286376719253505
16	83756670773733320287699303047996412235223138303
17	62707921196923889899446452602494921906963551482675201
18	9942119532215951589522891459235452451655502687858830501478 3
19	3327719012271075917361775733112611258835830762584219025835467735 05
20	2344880451051088988152559855229099188899081192234291298795803236068491263

Likelihood

Network: G



LungCan = 0(no), 1(yes)
Bronch = 0(no), 1(yes)
Dysponea = 0(no), 1(yes)

L = P(G|D) = ?

Observed data:

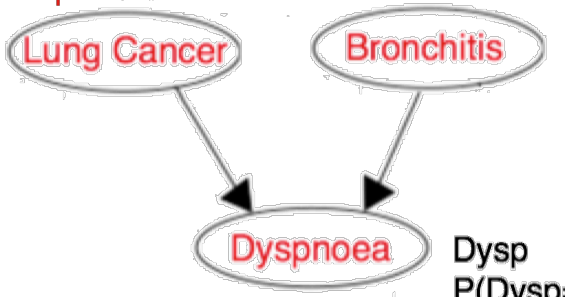
LungCan	Bronch	Dysp
1	0	1
0	1	1
0	1	1
0	1	0
0	1	1
0	0	0
0	1	1
0	1	1
0	1	1
1	1	1
0	1	1
0	0	0
0	1	1
0	0	0
1	0	1

Likelihood

Calculating conditional probability tables:

LungCan	x=1	x=0
P(LungCan=x)	1/5	4/5

Bronch	x=1	x=0
P(Bronch=x)	2/3	1/3



Dysp
P(Dysp=x|LungCan=y,Bronch=z)

y	z	x=1	x=0
0	0	0	1
0	1	8/9	1/9
1	0	1	0
1	1	1	0

Observed data:

LungCan	Bronch	Dysp
1	0	1
0	1	1
0	1	1
0	1	0
0	1	1
0	0	0
0	1	1
0	1	1
0	1	1
1	1	1
0	1	1
0	0	0
0	1	1
0	0	0
1	0	1

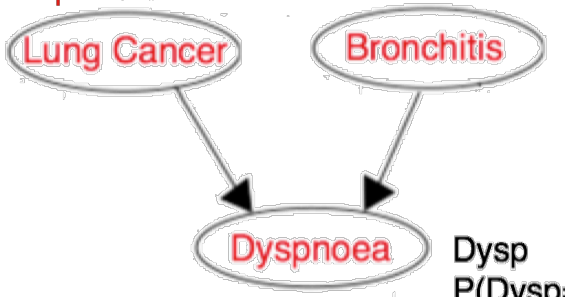
$$L = P(G|D) = \prod_{i=1}^{15} P(LungCan_i)P(Bronch_i)P(Dysp_i|LungCan_i, Bronch_i)$$

Likelihood

Calculating conditional probability tables:

LungCan	x=1	x=0
P(LungCan=x)	1/5	4/5

Bronch	x=1	x=0
P(Bronch=x)	2/3	1/3



Dysp
P(Dysp=x|LungCan=y,Bronch=z)

y	z	x=1	x=0
0	0	0	1
0	1	8/9	1/9
1	0	1	0
1	1	1	0

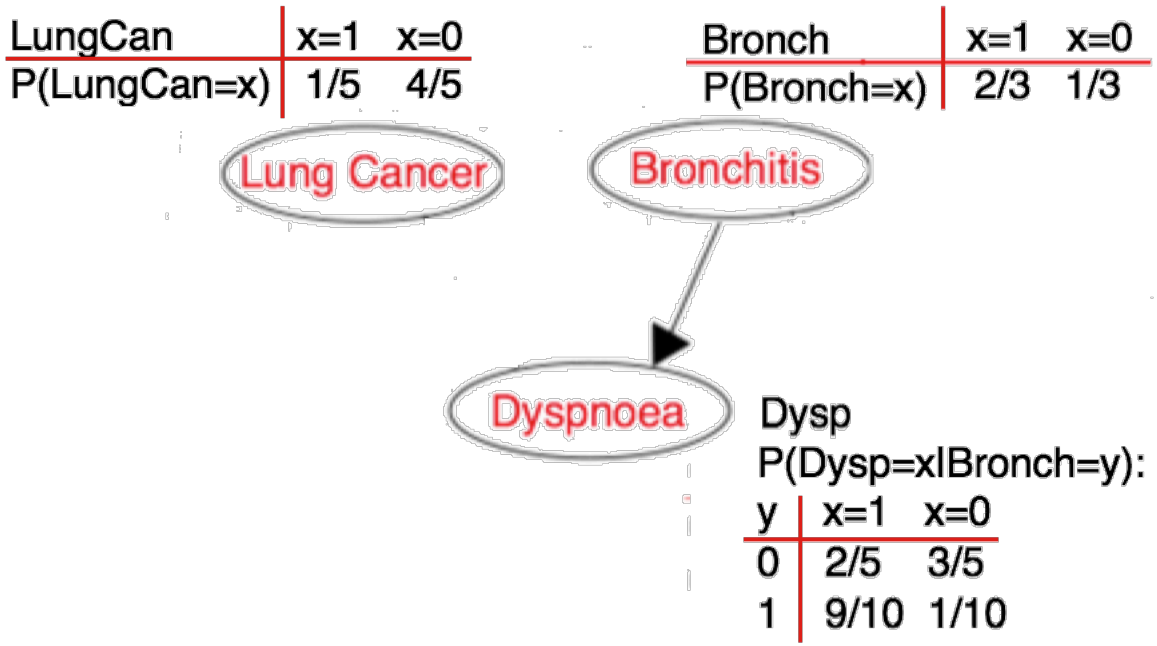
Observed data:

LungCan	Bronch	Dysp
1	0	1
0	1	1
0	1	1
0	1	0
0	1	1
0	0	0
0	1	1
0	1	1
0	1	1
1	1	1
0	1	1
0	0	0
0	1	1
0	0	0
1	0	1

L = P(G|D) = -20.1932377 (see R-code)

Likelihood

Network G_1



Observed data:

LungCan	Bronch	Dysp
1	0	1
0	1	1
0	1	1
0	1	0
0	1	1
0	0	0
0	1	1
0	1	1
0	1	1
1	1	1
0	1	1
0	0	0
0	1	1
0	0	0
1	0	1

$$L = P(G|D) = \prod_{i=1}^{15} P(LungCan_i)P(Bronch_i)P(Dysp_i|Bronch_i) = -23.6696369 \text{ (see R-code)}$$

BN structure learning

- **Constraint-based** algorithms #PC
  - start with a full graph
  - eliminate edges by performing conditional independence tests in a certain order
  - (!) prone to statistical testing errors
- **Search and Score** algorithms #GES (greedy equivalent search)
  - score function: log likelihood/posterior
  - search strategy is usually greedy due to the size of the search space
  - (!) local optima problem
- **MCMC (Markov chain Monte Carlo) schemes**
  - allow to obtain a sample from posterior distribution of graph given the data (rather than 1 single model)
  - avoid local optima problem
  - (!) convergence issues for structure MCMC

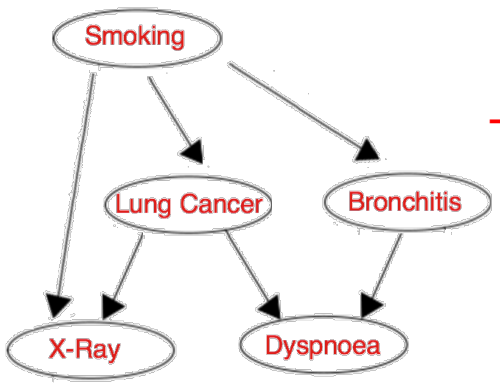
Learning BNs with R

R packages for learning BNs:

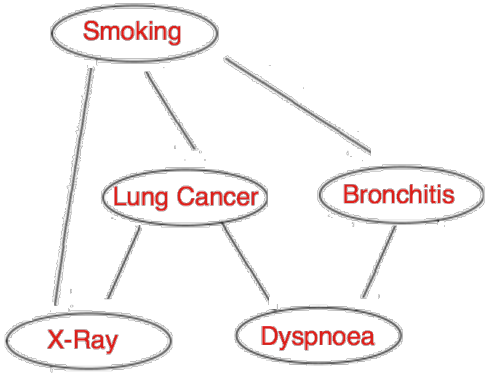
- pcalg – PC algorithm, greedy equivalent search (GES)
  https://cran.r-project.org/web/packages/pcalg/pcalg.pdf
- bnlearn – max-min hill climbing (MMHC), greedy search
  http://www.bnlearn.com/
- BiDAG – a collection of MCMC methods for MAP search and sampling from posterior
  https://cran.r-project.org/web/packages/BiDAG/BiDAG.pdf

Comparing BNs to each other

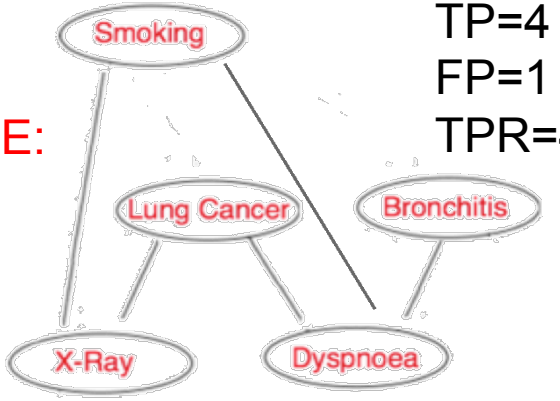
- Skeleton: TP, FP, TPR=TP/n.edges



T:



E:

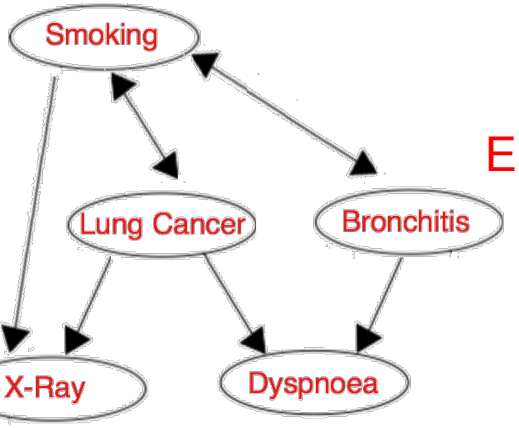


TP=4
FP=1
TPR=4/6=67%

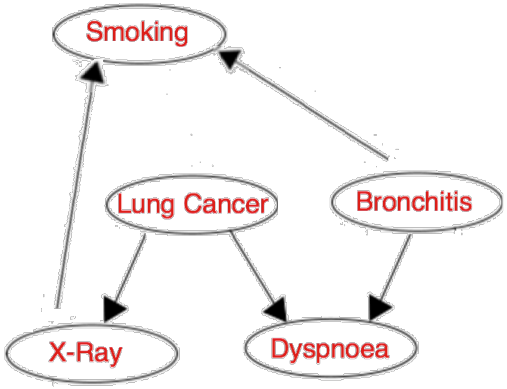
- CPDAG (equivalence class): SHD=FN+FP+EWD

  Where EWD are edges with wrong directions

T:



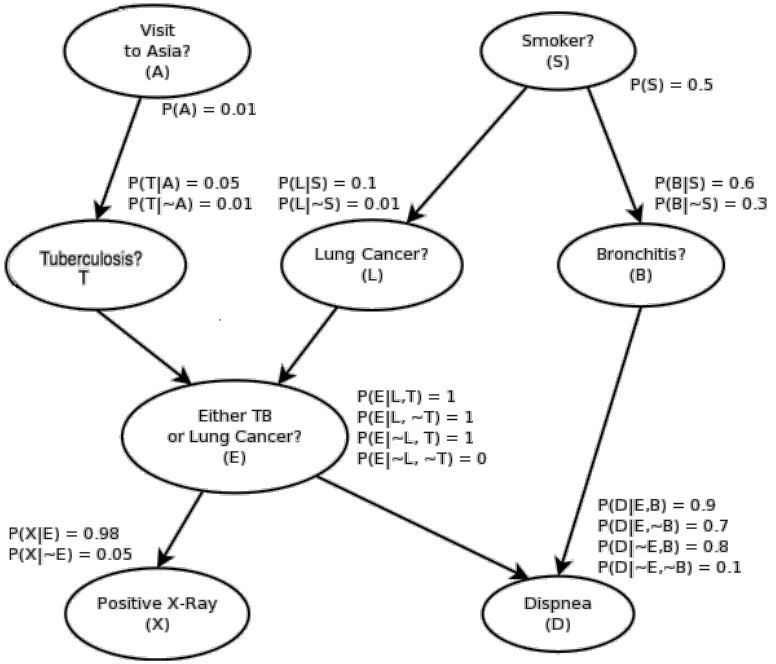
E:



SHD=3
TP=5
FP=0
TPR=5/6=83%

Asia dataset

- a synthetic dataset from Lauritzen and Spiegelhalter (1988)
- 8 binary variables:
  - `D` (*dyspnoea*), a two-level factor with levels `yes` and `no`.
  - `T` (*tuberculosis*), a two-level factor with levels `yes` and `no`.
  - `L` (*lung cancer*), a two-level factor with levels `yes` and `no`.
  - `B` (*bronchitis*), a two-level factor with levels `yes` and `no`.
  - `A` (*visit to Asia*), a two-level factor with levels `yes` and `no`.
  - `S` (*smoking*), a two-level factor with levels `yes` and `no`.
  - `X` (*chest X-ray*), a two-level factor with levels `yes` and `no`.
  - `E` (*tuberculosis versus lung cancer/bronchitis*), a two-level factor with levels `yes` and `no`.
- 5000 observations



http://www.bnlearn.com/documentation/man/asia.html

Asia dataset, + R-code

- Estimate equivalence class (CPDAG) from observational data using PC-algorithms and MCMC iterative scheme
- What are the differences between estimates of 2 algorithms?
- Compare estimated structures to the true DAG
- Why was edge A->T missed by both algorithms?
- Try the same with 500 observations instead of 5000
- Did the results change?

n=100, see R-code

- Generate random DAG with 100 nodes and data
- Apply PC and MCMC algorithms for structure learning
- Compare estimates to the true structure
- Is there a significant difference between results of 2 algorithms?
- Is there a significant difference between MAP estimate and sampling version of MCMC schemes?

References

- Chickering, D. M. (2002). Optimal structure identification with greedy search. Journal of Machine Learning Research 3, 507–554.
- Friedman, N. and D. Koller (2003). A Bayesian approach to structure discovery in bayesian networks. Machine Learning 50, 95–125.
- Kalisch, M. and P. Bühlmann (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. Journal of Machine Learning Research 8, 613–636.
- Kalisch, M., M. Mächler, D. Colombo, M. Maathuis, and P. Bühlmann (2012). Causal inference using graphical models with the R package pcalg. Journal of Statistical Software 47, 1–26.