BN Construction

- Identify variables
- Order them*
- While there are variables to add
  - Pick the next in the ordering
  - Identify its parents in the net
    - Hold all others constant (in every configuration)
    - If net variable can influence it, that net var is a parent
  - Draw all arcs and add CPT

* order matters a lot

Heuristic

Usually the most compact representation results when belief causality mirrors physical causality

Dentist Example

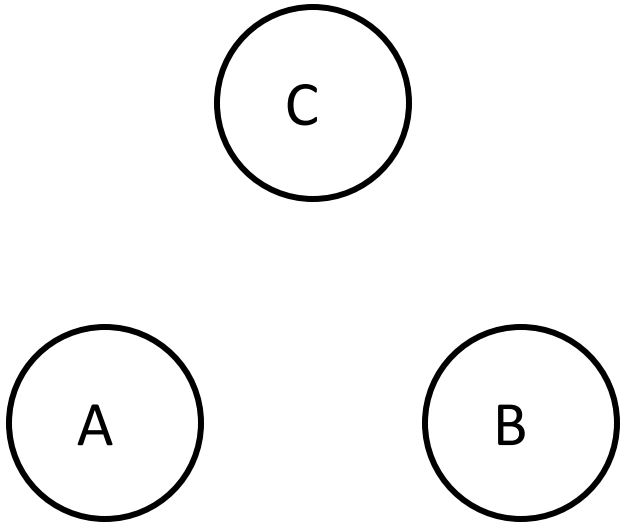
3 Boolean Random Variables:

C – Patient has a cavity

A – Patient reports a toothache

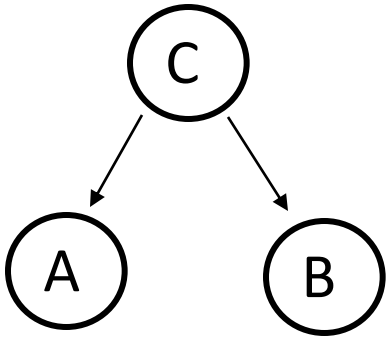
B – Dentist's probe catches on tooth

Joint: 7 parameters



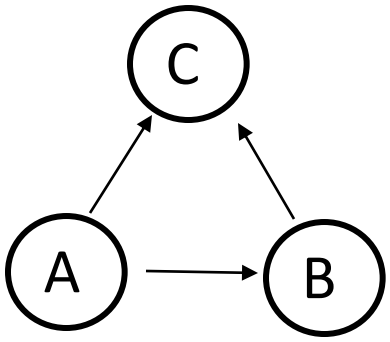
Effect of Order on BN structure

Order: C A B



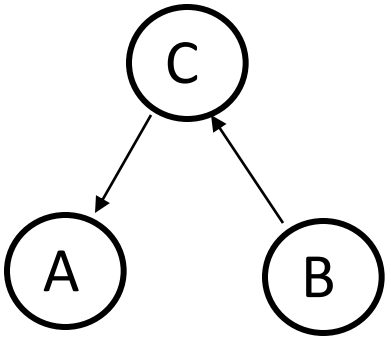
5 parameters

Order: A B C



7 parameters
no savings over
the joint

Order: B C A



5 parameters

BN Structure

consider ordering M,J,E,B,A

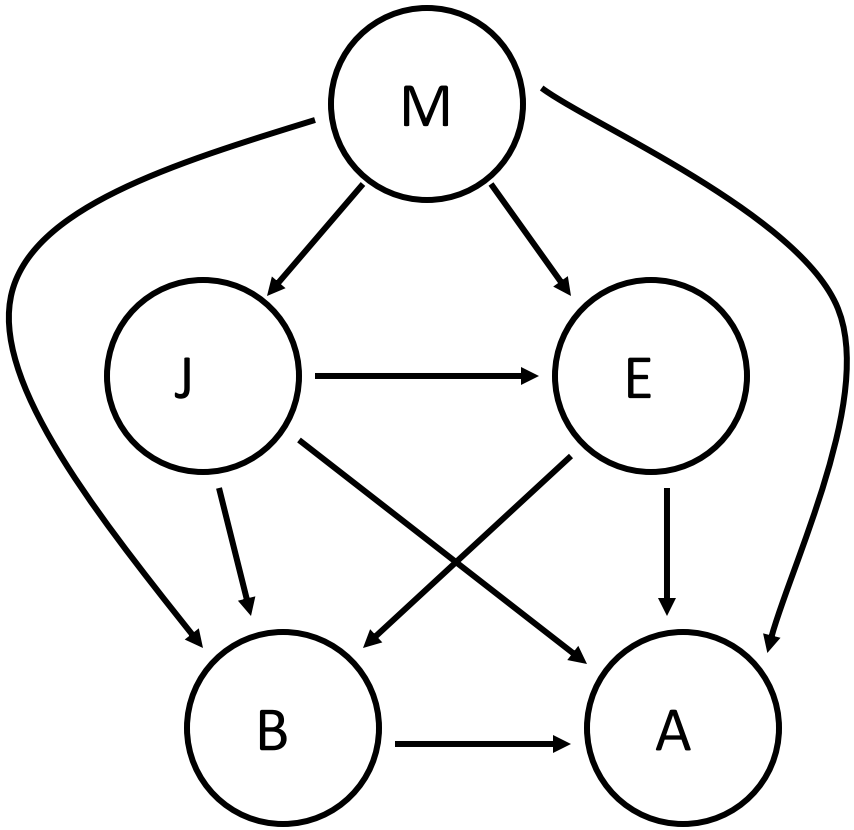
B – a burglary is in progress

E – an earthquake is in progress

A – the alarm is sounding

J – John calls

M – Mary calls



All variables are Boolean

How many numbers?

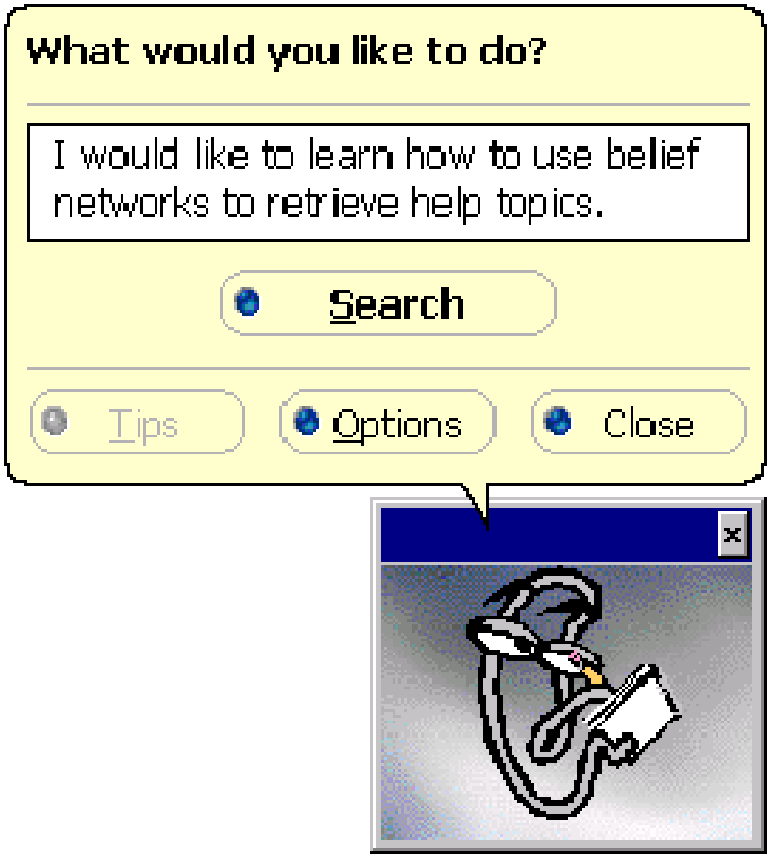
No savings over the Joint

What conditional independence assumptions does it make?

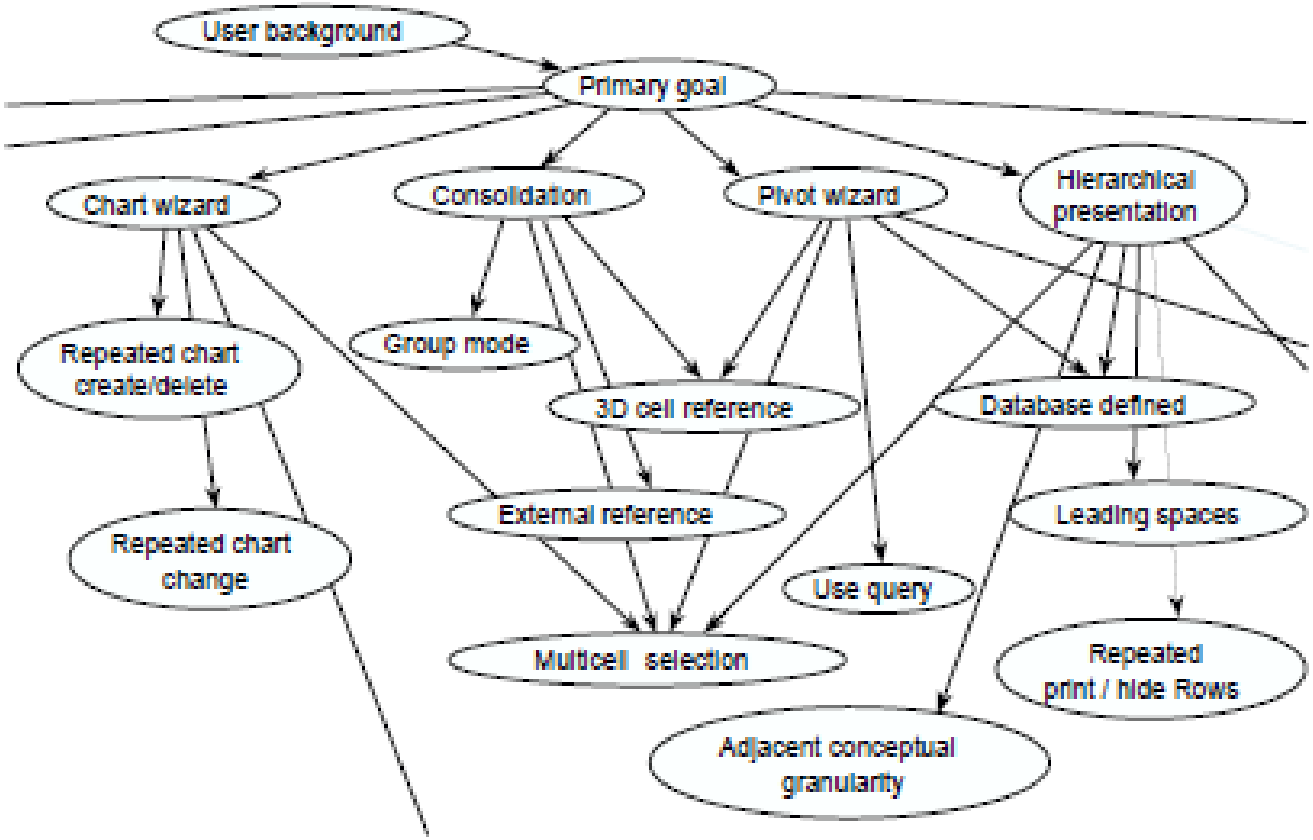
Ordering is Important

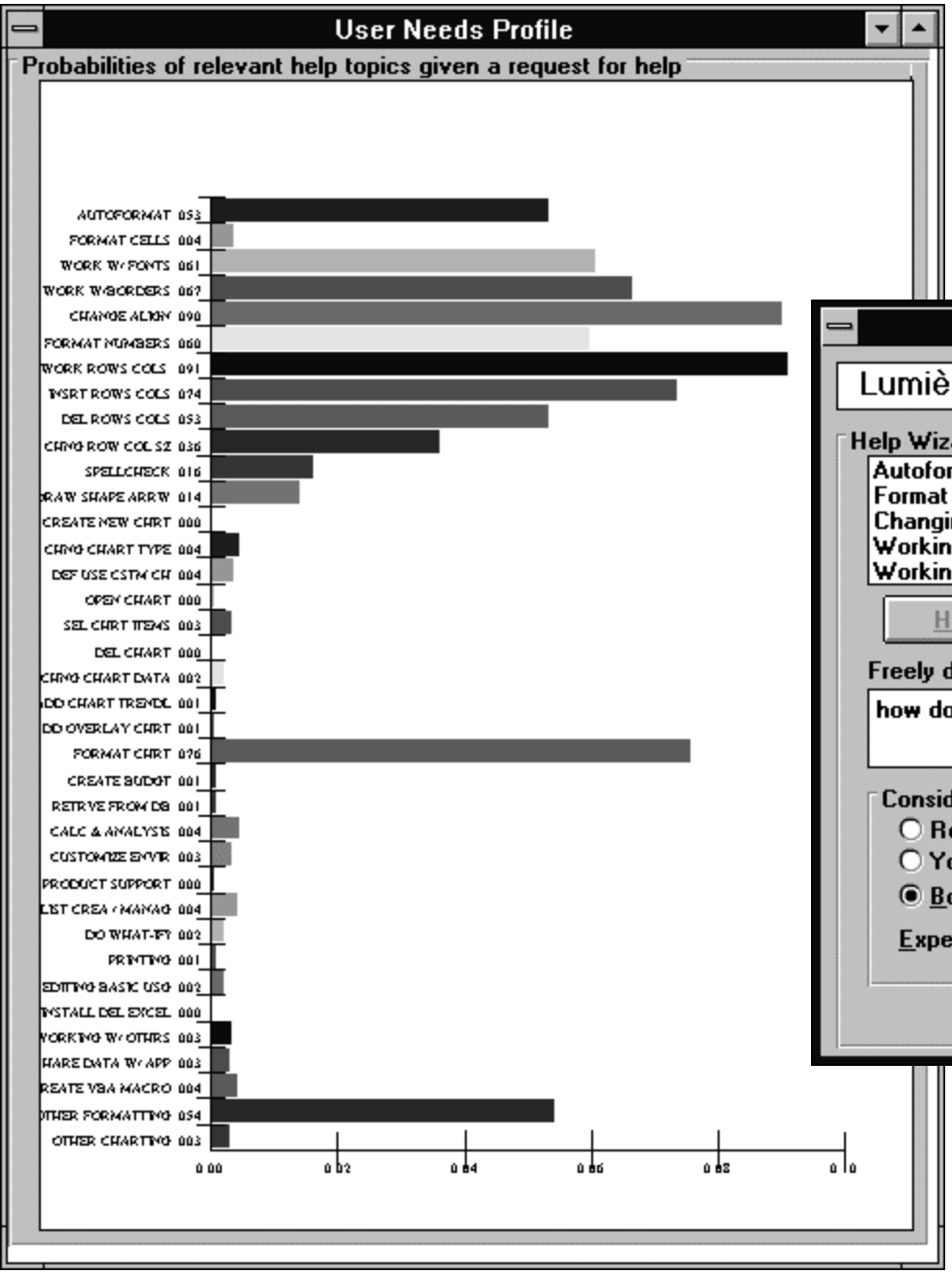
- Original net B E A J M
  - 10 parameters
  - Saves 21 over the joint
- Ordering M J E B A
  - 31 parameters
  - No savings over the joint
- “Causality” ordering is only heuristic
- Suppose we want to test all orderings of n random variables…

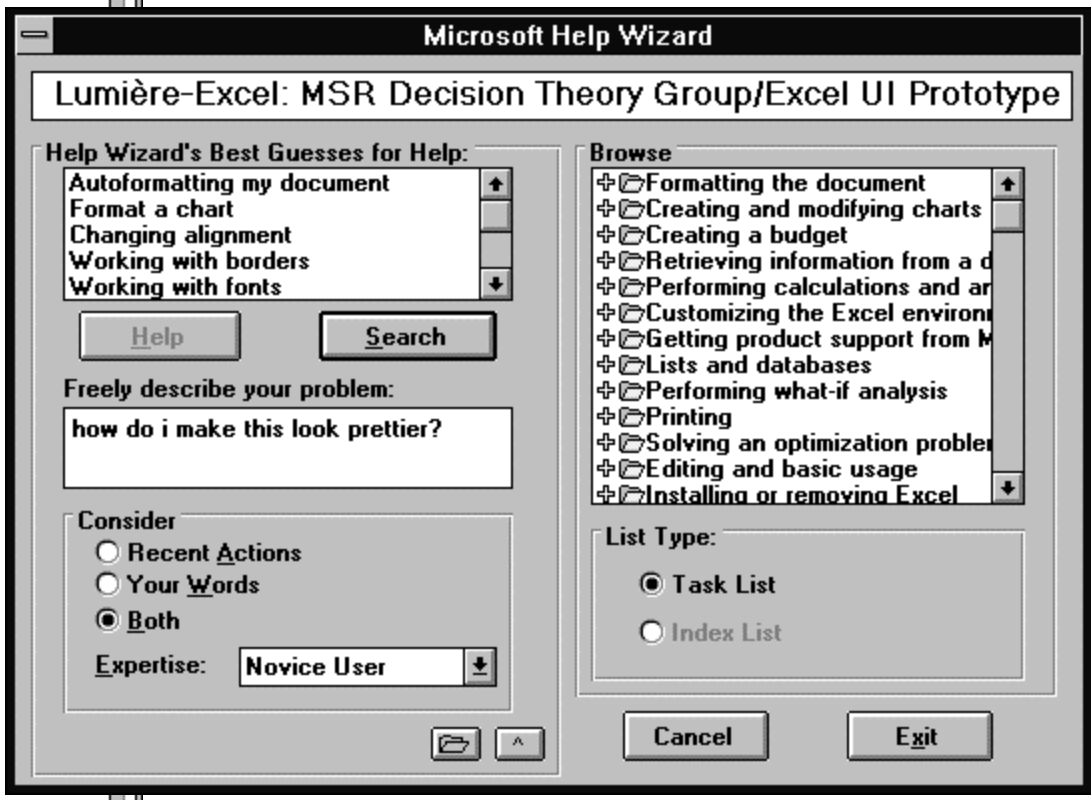
Clippy (Lumiere Project): Another Real World(?) BN Example

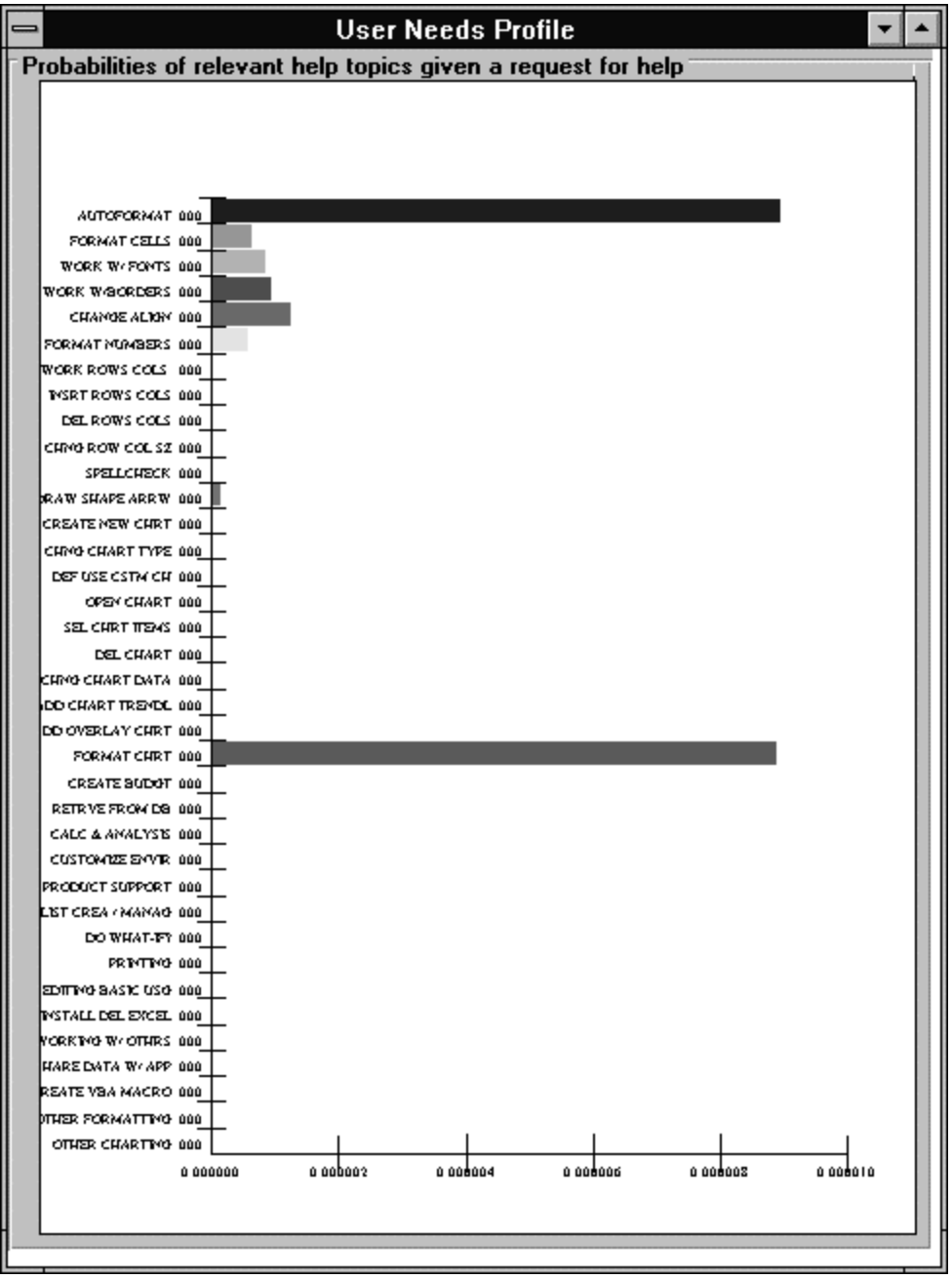


Portion of Clippy's BN for Excel User Modeling




User Needs Profile
Probabilities of relevant help topics given a request for help
AUTOFORMAT 053
FORMAT CELLS 004
WORK W/ FONTS 061
WORK W/BORDERS 067
CHANGE ALIGN 090
FORMAT NUMBERS 060
WORK ROWS COLS 091
INSRT ROWS COLS 074
DEL ROWS COLS 053
CHNG ROW COL SZ 036
SPELLCHECK 016
DRAW SHAPE ARRW 014
CREATE NEW CHRT 000
CHNG CHART TYPE 004
DEF USE CSTM CH 004
OPEN CHART 000
SEL CHRT ITEMS 003
DEL CHART 000
CHNG CHART DATA 002
ADD CHART TRENDL 001
ADD OVERLAY CHRT 001
FORMAT CHRT 076
CREATE BUDGT 001
RETRVE FROM DB 001
CALC & ANALYSIS 004
CUSTOMIZE ENVIR 003
PRODUCT SUPPORT 000
LIST CREA/MANAG 004
DO WHAT-IF? 002
PRINTING 001
EDITING BASIC USG 002
INSTALL DEL EXCEL 000
WORKING W/ OTHRS 003
SHARE DATA W/ APP 003
CREATE VBA MACRO 004
OTHER FORMATTING 054
OTHER CHARTING 003
0.00
0.02
0.04
0.06
0.08
0.10


Microsoft Help Wizard
Lumière-Excel: MSR Decision Theory Group/Excel UI Prototype
Help Wizard's Best Guesses for Help:
Autoformatting my document
Format a chart
Changing alignment
Working with borders
Working with fonts
Help
Search
Freely describe your problem:
how do i make this look prettier?
Consider
Recent Actions
Your Words
Both
Expertise:
Novice User
Browse
Formatting the document
Creating and modifying charts
Creating a budget
Retrieving information from a d
Performing calculations and ar
Customizing the Excel environ
Getting product support from M
Lists and databases
Performing what-if analysis
Printing
Solving an optimization proble
Editing and basic usage
Installing or removing Excel
List Type:
Task List
Index List
Cancel
Exit



Posterior

- Query nodes after
- Conditioning
- Marginalizing

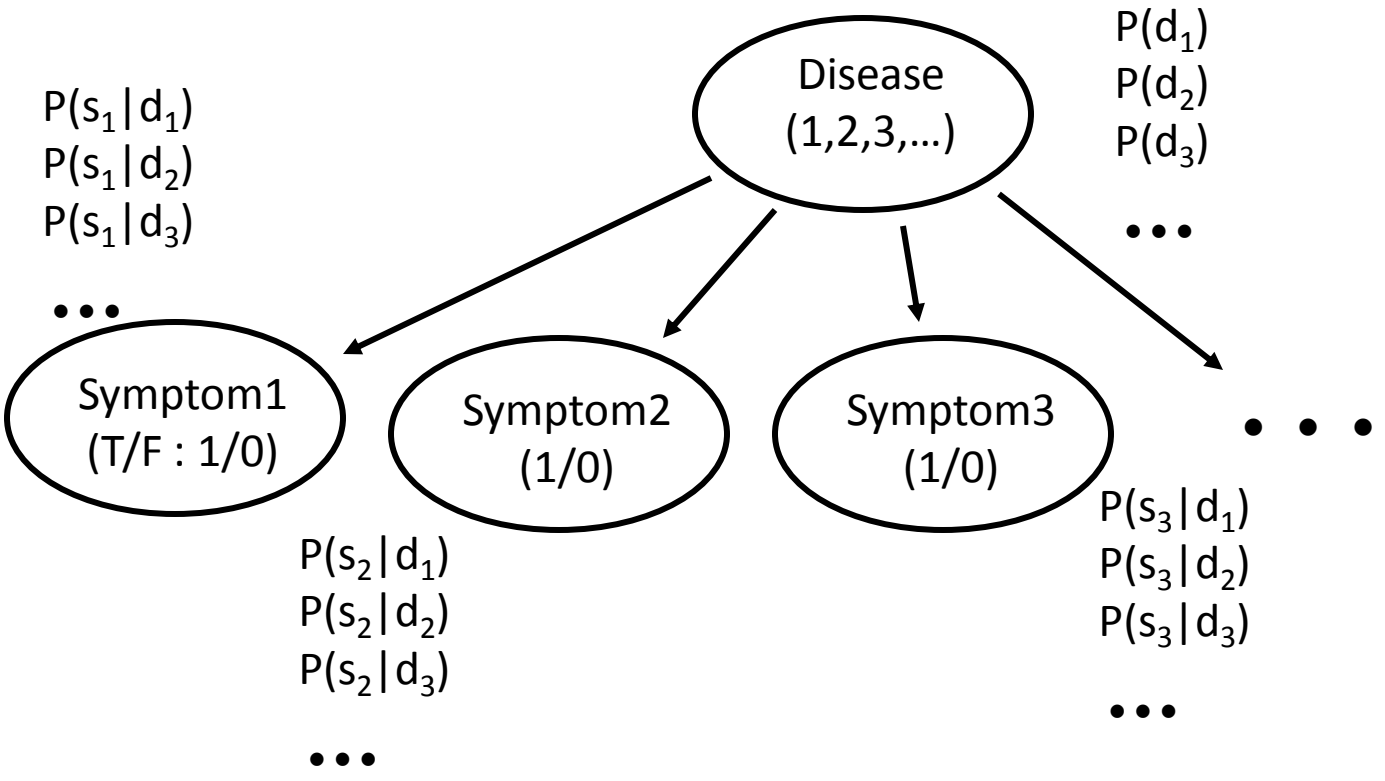
Naïve Bayes: a Simple and Important BN type

- Assume Boolean Attributes (Symptoms) are conditionally independent of each other given the Classification (Disease)
- Impoverished expressiveness
- Works surprisingly (?) well in practice
- Common first choice

NOTE the use of intentionally impoverished model for computational convenience

What does a naïve BN look like?

Naïve Bayes



Infer most likely disease:

$$\arg\max_{d_i} P(d_i \mid \overline{S})$$

Naïve Bayes

$$\arg\max_{d_i} P(d_i \mid \overline{S})$$

$$\arg\max_{d_i} \frac{P(d_i) \cdot P(\overline{S} \mid d_i)}{P(\overline{S})}$$

$$\arg\max_{d_i} \frac{P(d_i) \cdot \prod_j P(s_j \mid d_i)}{P(\overline{S})}$$

$$\arg\max_{d_i} \left(P(d_i) \prod_j P(s_j \mid d_i) \right)$$

Diagnose by inference with:

First three are probability models

|D| Functions, each assigns probability over the Boolean hypercube

If the probability models were correct we would no longer need to see the world

Fourth is not normalized (why does that work?)

Parameters are adjusted to best fit the world training samples

Fitting models to data is a kind of machine learning

Naïve Bayes

$$\arg\max_{d_i}\left(P(d_i)\prod_j P(s_j \mid d_i)\right)$$

- Suppose we always reason from observed symptoms to diseases & symptoms are Boolean
- Characterize the decision boundaries
- Log(x) is monotonically increasing, so:

$$\arg\max_{d_i} Log\left(P(d_i)\prod_j P(s_j \mid d_i)\right)$$

- Log of a product is…

Naïve Bayes

$$\arg\max_{d_i} Log\left(P(d_i)\prod_j P(s_j \mid d_i) \right)$$

- We represent the symptoms S (evidence) is a Boolean vector:

  s_k = 1 if k'th symptom is present
  = 0 if absent

- Becomes argmax $_i$ of $F_i(S)$
  where S is the Boolean vector of evidence

- $F_i(S) = \text{Log } P(d_i) + \sum \text{Log } P(s_k \mid d_i)$

- Form is $F_i(S) = a_i + B_i \cdot S$

  $a_i = \text{Log } P(d_i) + \sum \text{Log } P(s_k=0 \mid d_i)$
  $b_{ik} = \text{Log } P(s_k=1 \mid d_i) - \text{Log } P(s_k=0 \mid d_i)$

- The Log Prob fcns are…

  hyperplanes cutting through the Boolean hypercube

Naïve Bayes

$$\arg\max_{d_i} Log\left(P(d_i) \prod_j P(s_j \mid d_i) \right)$$

- Form $F_i(S) = a_i + B_i \cdot S$
  Each is a linear polynomial in S
- Diagnose using the highest valued function at a point S
- What are the Naïve Bayes decision boundaries?
- (How do hyperplanes interact?)
- How do we determine a's and B's?
- These are *generative models* for the diseases