Learning Analytics

Dr. Bowen Hui
Computer Science
University of British Columbia Okanagan

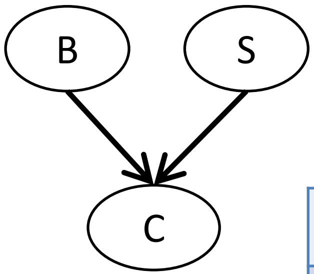
Matlab BNT

- Download at: https://github.com/bayesnet/bnt
- Documentation at: https://www.cs.utah.edu/~tch/notes/matlab/bnt/docs/bnt_pre_sf.html
- "How to use the Bayes Net Toolbox"
  - Installation
    - Download, unzip, start Matlab, cd to path containing files, run:
      - addpath(genpathKPM(pwd))
      - test_BNT
    - After exiting Matlab, need to addpath at start of next use
  - Creating your first Bayes net

Recall BN Example #4

- Pr(B=true | C=true, S=true) = ?
- Pr(B=true | C=true) = ?

Pr(~B)	Pr(B)
0.5	0.5



Pr(~S)	Pr(S)
0.5	0.5

		Pr(~C|B,S)	Pr(C|B,S)
~B	~S	1	0
B	~S	0	1
~B	S	0	1
B	S	0	1

Implementing BN Example #4

```
B = 1; S = 2; C = 3;
dag = zeros(3,3);                 % directed acyclic graph
dag([B S], C)=1;
ns = 2*ones(1,3);
bnet = mk_bnet(dag, ns);          % makes the BN structure

% populate CPTs
bnet.CPD{B} = tabular_CPD(bnet, B, 'CPT', [0.5 0.5]);
bnet.CPD{S} = tabular_CPD(bnet, S, 'CPT', [0.5 0.5]);
CPT = zeros(2,2,2);
CPT(1,1,:) = [1 0];          % make sure each row sums up to 1.0
CPT(2,1,:) = [0 1];
CPT(1,2,:) = [0 1];
CPT(2,2,:) = [0 1];
bnet.CPD{C} = tabular_CPD(bnet, C, 'CPT', CPT);
```

Inference with BN Example #4

```
% matlab convention: 1=false 2=true
engine = jtree_inf_engine(bnet);
ev = cell(1,3);

ev{C} = 2;    % C is observed to be true
engine = enter_evidence(engine, ev);
m = marginal_nodes(engine, B);
fprintf('P(B=true|C=true) = %5.3f\n', m.T(2))

ev{S} = 2;    % C and S are now true
engine = enter_evidence(engine, ev);
m = marginal_nodes(engine, B);
fprintf('P(B=true|C=true,S=true) = %5.3f\n', m.T(2))
```

General Guidelines for Building BNs

- We usually do not build the model based on knowledge about the joint probability distribution

- Typically, we have some vague idea of the dependencies in the world, then we define it precisely into a graphical model

- Steps to follow:
  - Formulate the problem
  - Define the RVs involved
  - Choose independence relations
  - Assign probabilities in the CPTs

Guidelines for Choosing RVs

- Variables must be precise
  - What are the values?
  - How to define them?
  - How to measure them?
  - E.g. weather: difference between the values cold vs. bitter-cold?

- Our discussion: discrete variables

- Different kinds of variables:
  - Observable
  - Hidden – may or may not be useful to include, depending on other independencies they generate

Guidelines for Building the Graph

- When we have information about causality, use causal connections to simplify graph

- Consider tradeoffs between precision of the model and size/sparsity of the graph

Guidelines for Defining Numerical Parameters

- Where do the probabilities come from?
  - Expert
  - Approximate analysis
  - Guessing
  - Learning from data

Guidelines for Defining Numerical Parameters

- Where do the probabilities come from?
  - Expert
  - Approximate analysis
  - Guessing
  - Learning from data

- **Bad news:**
  - In all these cases, the numbers are approximate!

Guidelines for Defining Numerical Parameters

- Where do the probabilities come from?
  - Expert
  - Approximate analysis
  - Guessing
  - Learning from data

- **Bad news:**
  - In all these cases, the numbers are approximate!
- **Good news:**
  - The numbers usually do not matter all that much
  - Sensitivity analysis can help decide if certain numbers are critical or not for the conclusions

Guidelines for Defining Numerical Parameters

- Avoid assigning zero probability to any event
- The relative values of conditional probabilities for $Pr(X_i | Par(X_i))$ given different values of $Par(X_i)$ is important
- Having probabilities that are orders of magnitude different can cause problems in the network

Key Ideas

- Main concept
  - Modeling guidelines for choosing RVs and defining causal relationships
  - Sensitivity analysis can help determine meaningful numeric parameters in the BN
- Computation:
  - Matlab BNT