Bayesian Networks
Basic and simple graphs

Ullrika Sahlin, Centre of Environmental and Climate Research

Lund University, Sweden

Ullrika.Sahlin@cec.lu.se

http://www.cec.lu.se/ullrika-sahlin

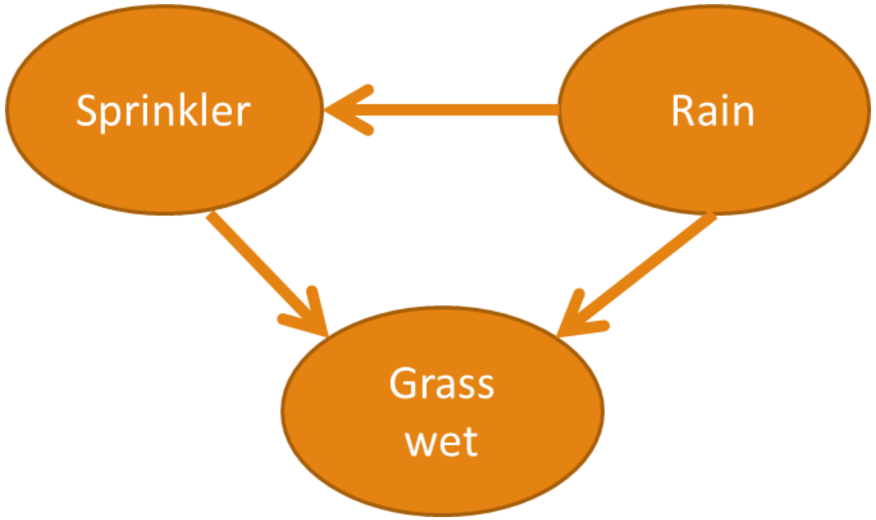
Bayesian [Belief] Networks

- A BN is a graphical model (graph) with nodes representing random variables and edges representing probabilistic dependencies.
- BNs enable us to model uncertain events.
- BNs provide an intuitive visual representation of assumptions or reasoning hidden in the head of an expert.
- BNs allow us to apply the laws of probability and Bayes rule to propagate consistently the impact of evidence on the probabilities of uncertain outcomes (i.e. to revise the probabilities in light of evidence).
- A BN can be made as an influence diagram which depicts the logical or causal relations among factors that influence the likelihood of outcome states of some parameter(s) of interest.
- A BN can be used to find Bayes optimal decisions

LUND UNIVERSITY

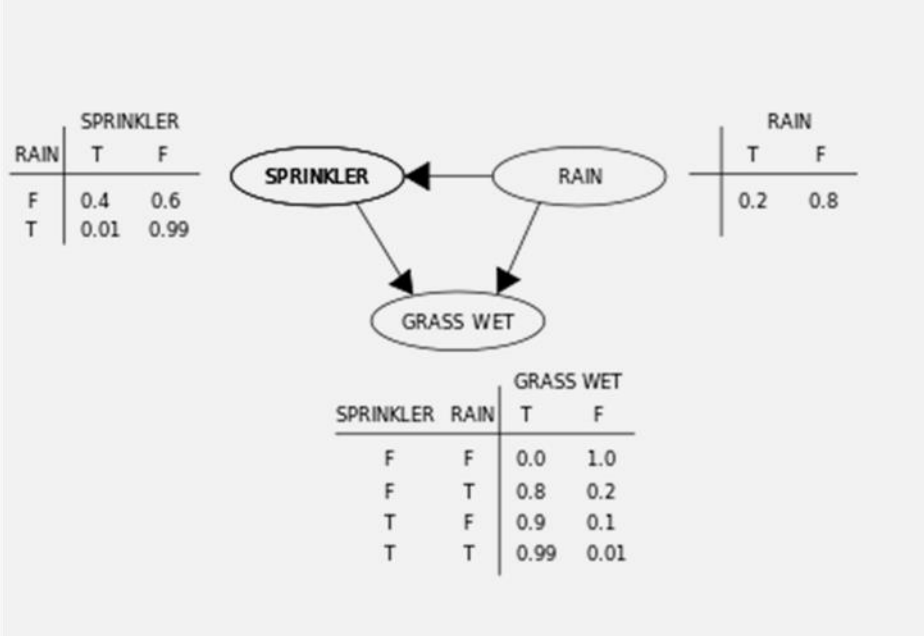
LUND UNIVERSITY

A BN is a graph



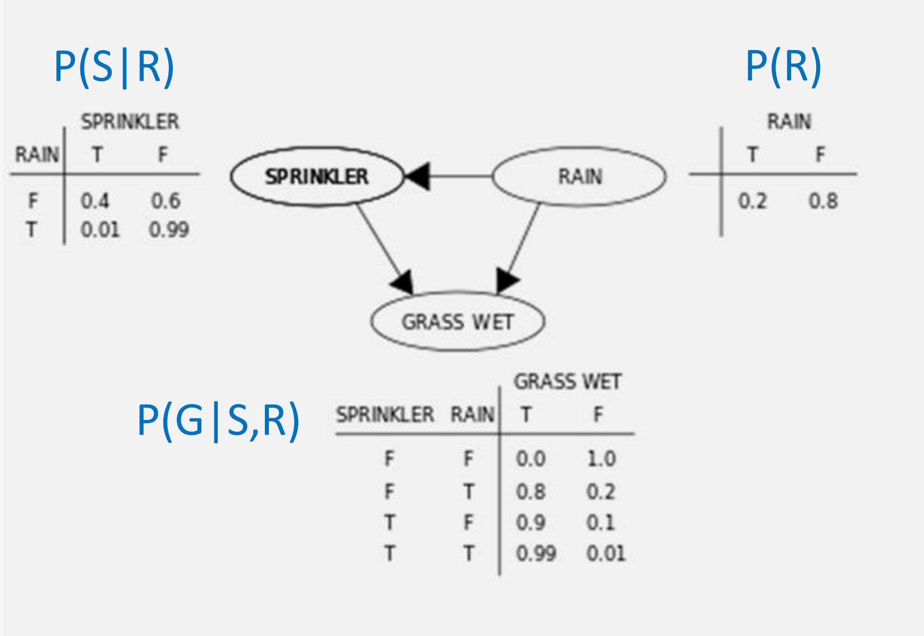
- Let X be a set of random variables
- A probabilistic graphical model consists of:
- Nodes V corresponding to variables in X
- Edges E revealing conditional independencies
- A graph G = (V,E)
- Probability distributions over V and E: ψ
- Parameters: θ

A BN is a graph with probabilistic links



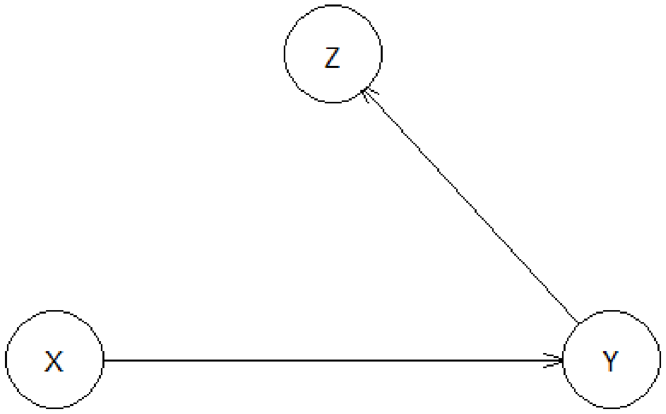
- A graphical probability model for P(SPRINKLER, RAIN, GRASS WET)
- Nodes V = {S, R, G}
- Edges E = {P(R), P(S|R), P(G|S,R)}
- Parameters θ = ”the elements in the probability tables”

A BN is a Directed Acyclic Graph

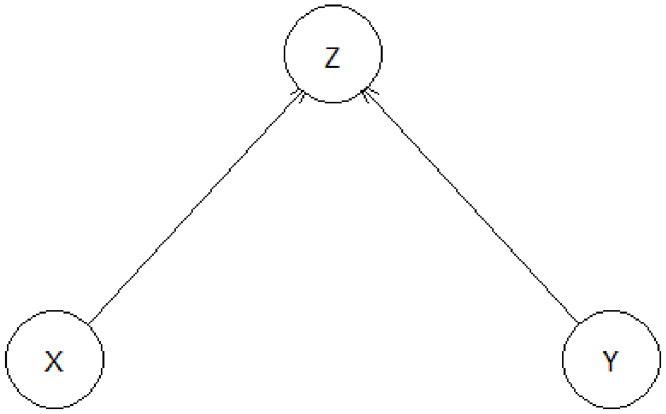


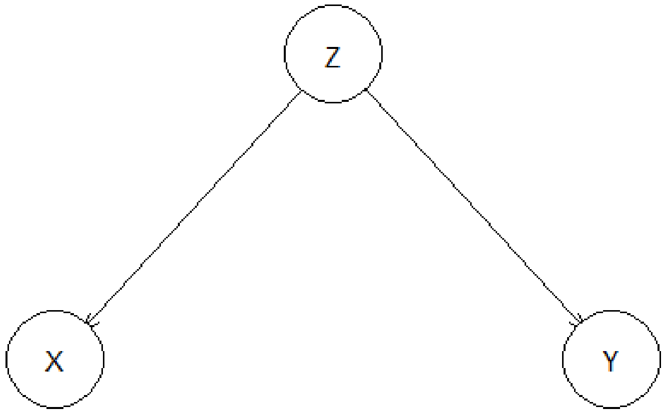
- The DAG express causal relations
- The DAG helps us to decompose the joint probability distribution of all nodes together
- $P(G,S,R) = P(G|S,R) \cdot P(S|R) \cdot P(R)$

Serial connection



Converging and divering connections





The marbled crayfish in Arbogaån

- Nov 2012: Individuals found in the Arboga river
- Seven individuals were removed from the site
- Winter 12/13 was very cold
- Removing any individuals left 2013 is urgent!
- Trial fishing performed during summer 2013 did not find any individuals
- Is the crayfish still present in the system?
- What should be done?


Aqua reports 2013:17

Marmorkräftan i Märstaån

Riskanalys och åtgärdsförslag

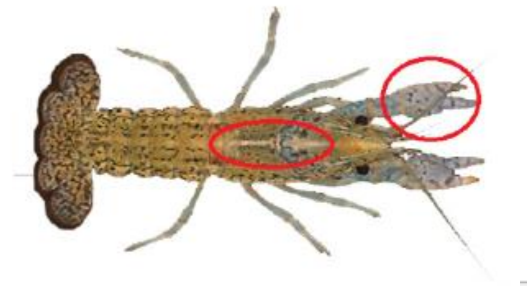
Patrik Bohman & Lennart Edsman

SLU

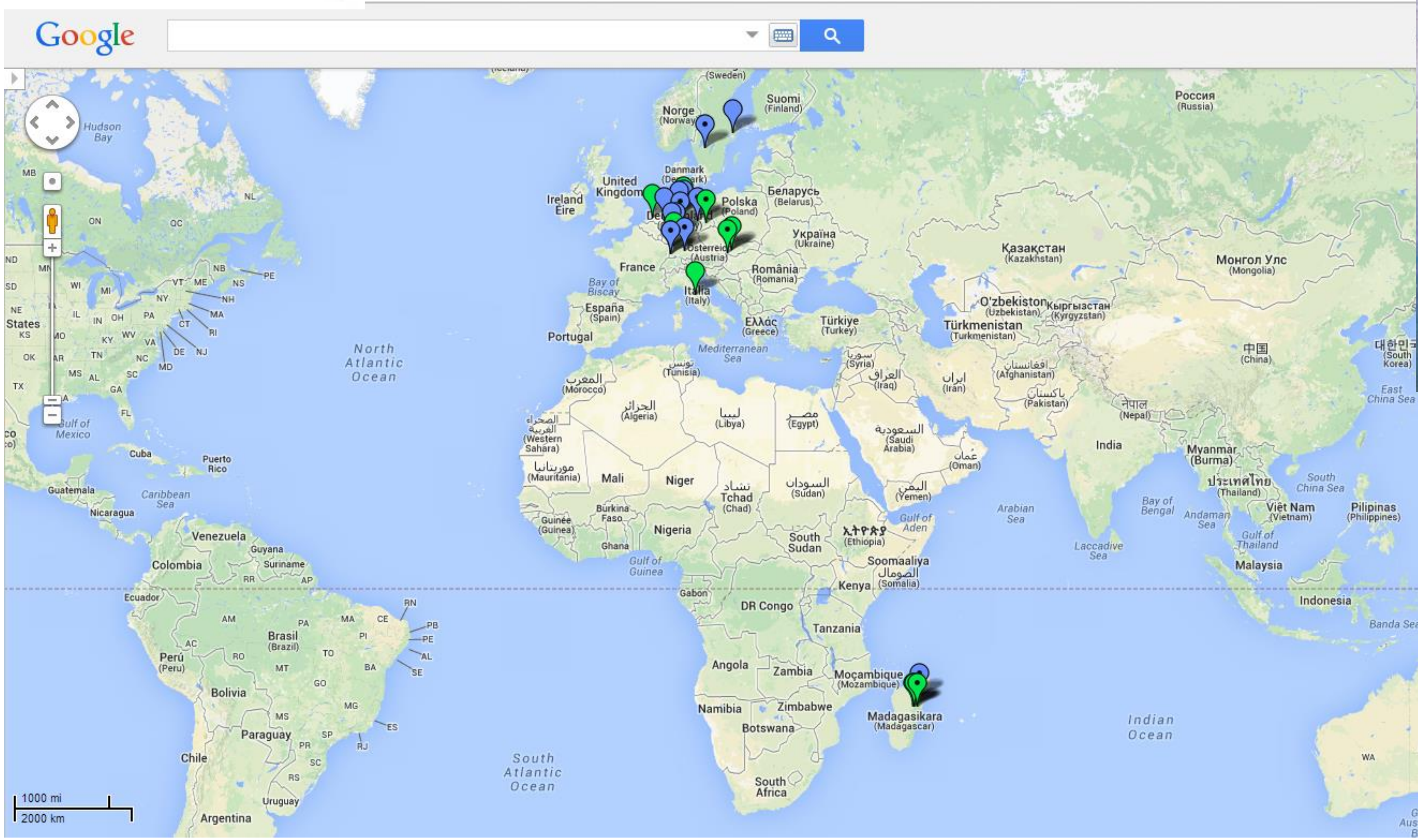
Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Institutionen för akvatiska resurser

Lund University

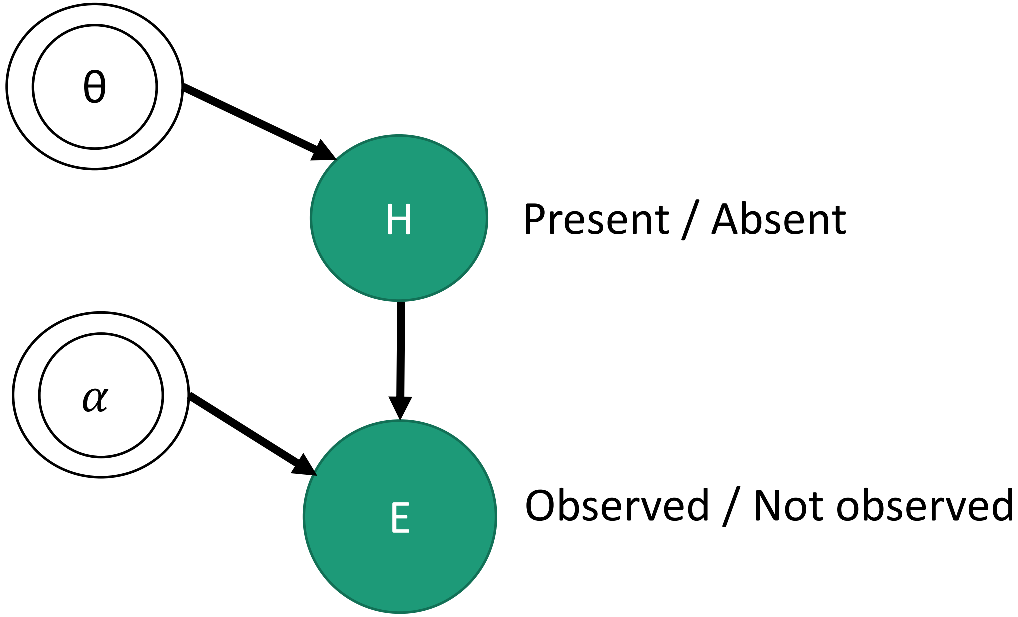


We don't want it in the wild!



LUND UNIVERSITY

Is the crayfish still present?



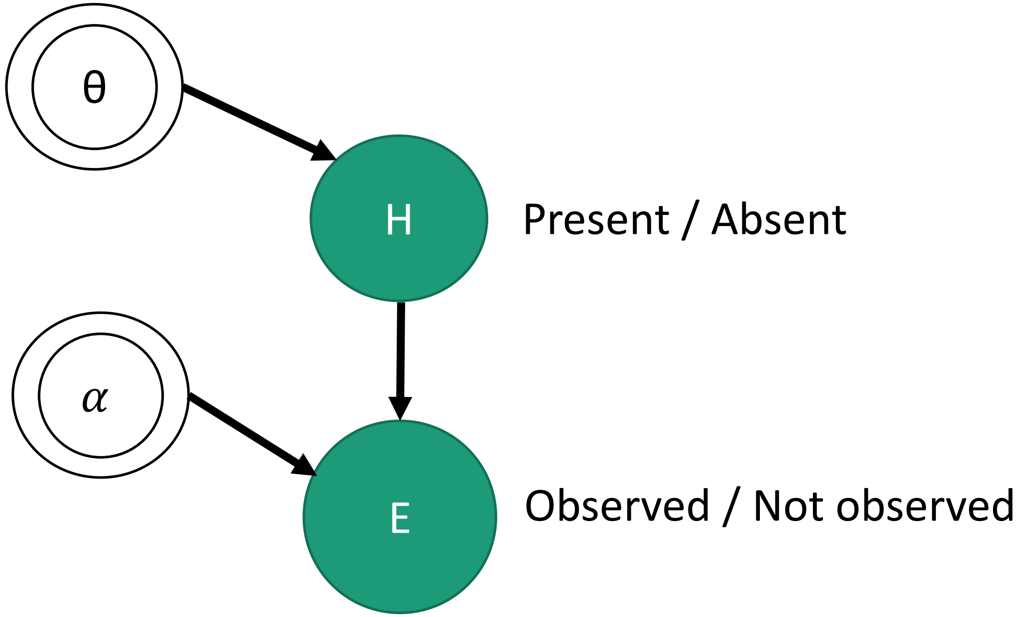
H	
Present	θ
Absent	1 − θ

	H	
E	Present	Absent
Observed	α	0
Not observed	$1-\alpha$	1

Is the crayfish still present?

P(H) = 0.23

P(E|H) = 0.3



H	
Present	θ
Absent	1 – θ

	H	
E	Present	Absent
Observed	α	0
Not observed	$1-\alpha$	1

$$P(H|not\ E) = \frac{(1\ -\alpha)\theta}{1-\alpha\theta}$$

P(H|not E) = 0.15

Build a network, update belief, generate a sample of predictions

- Draw the nodes
- Define the state of nodes
- Draw the directed edges (arrows)
- Populate the probability tables (add values, sum to one, elicitation tools)
- Configure the graph by instantiating nodes with hard evidence
- Update beliefs (F5)
- Note the belief (marginal probability distribution) of endpoint (what to predict)
- Generate data file – sample from the marginal, bias samples on existing evidence

A BN help us to revise our belief in light of new evidence

- Hard Evidence (instantiation) for node X is when the state of node X is definitely known
- Soft Evidence for node X is any evidence that enables us to update the prior probability values for the states of X
- Evidence is implemented as a configuration of the graph

- What is the probability that the crayfish is present if we did not observe any individuals?

LUND UNIVERSITY

BNs can be large networks

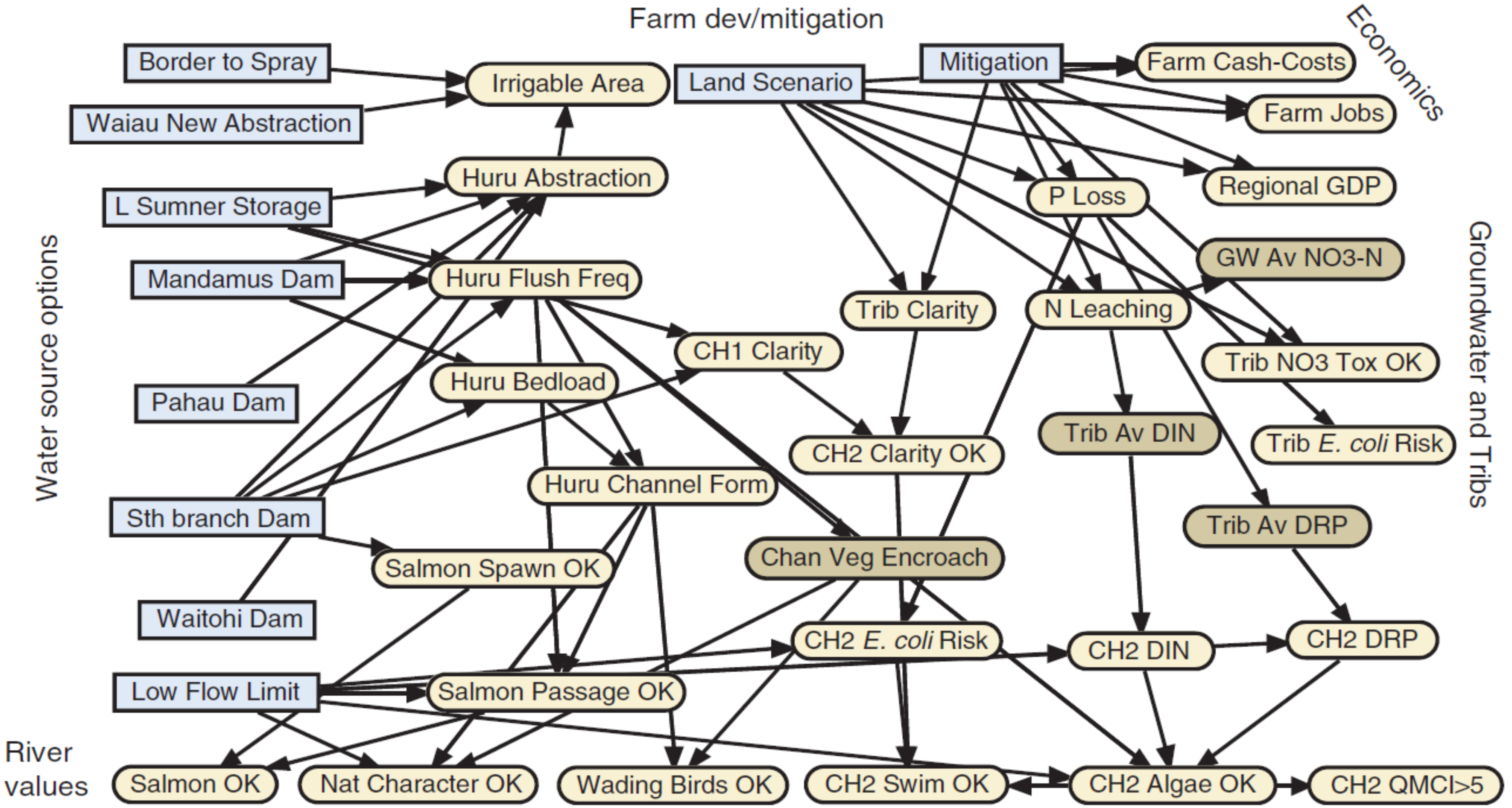
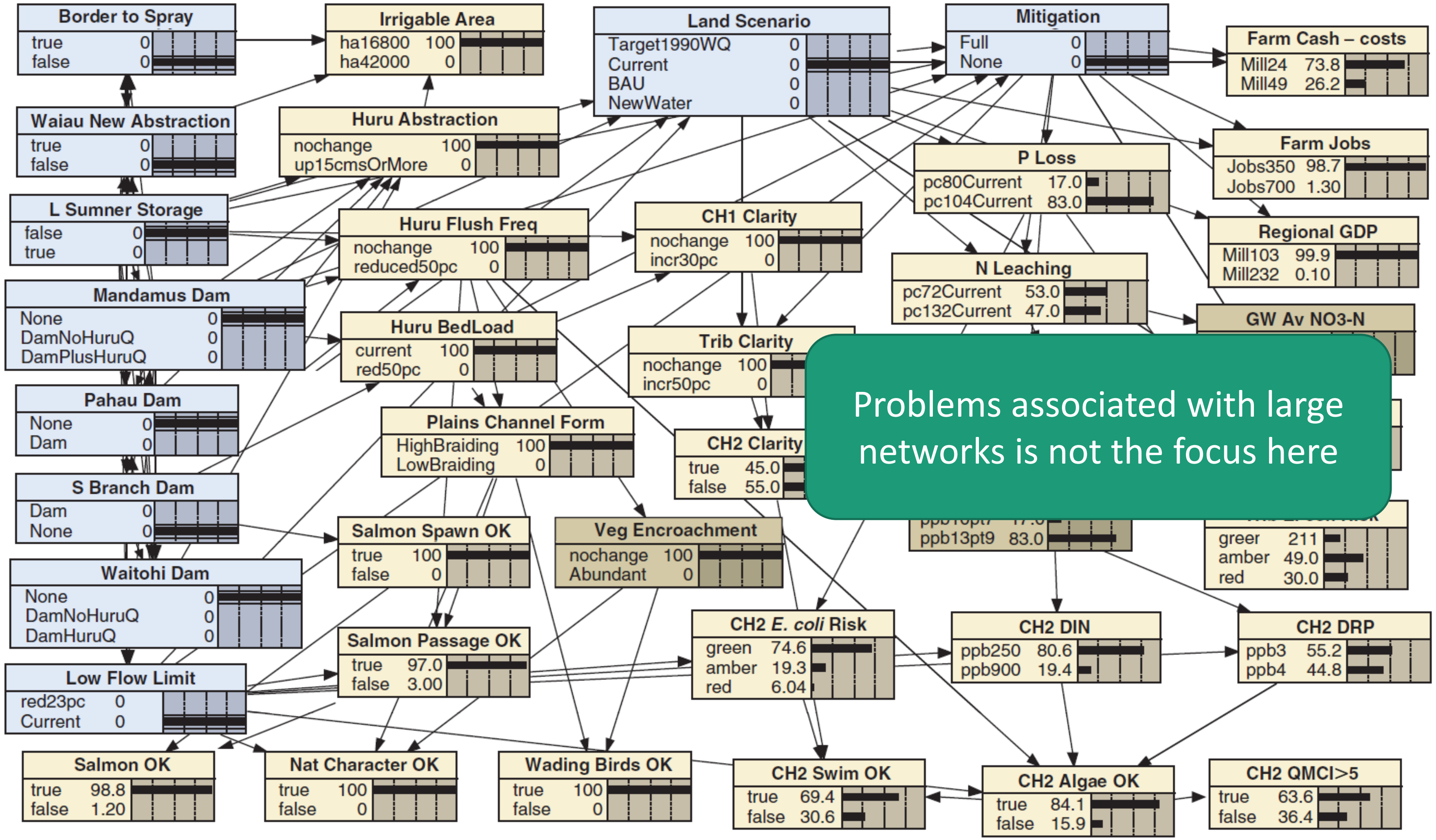


Fig. 2. Conceptual model of linkages between water-source and farm-development options, and stakeholder socioeconomic and environmental values. Rectangles (BN decision nodes) represent management actions and attributes in ellipses (BN nature nodes) respond to changes in actions.

LUND UNIVERSITY

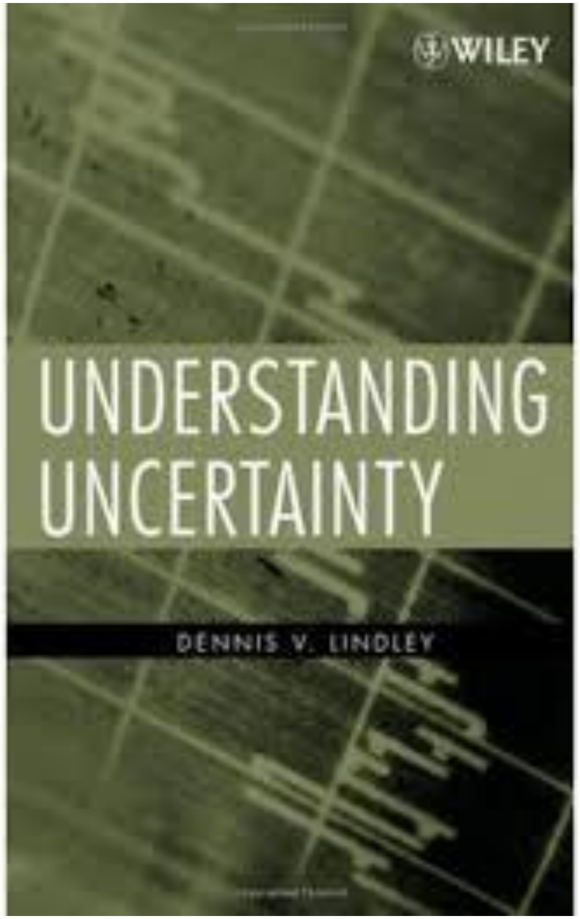
BNs can be large networks



Problems associated with large networks is not the focus here

Bayesian principles to learn and make predictions

A recapitulation



Conditioning

- K is our knowledge bases right now*
- P(A|K)
- E is a new piece of evidence, e.g. a weather forecast
- P(A|E & K)

* The model is also part of the knowledge bases

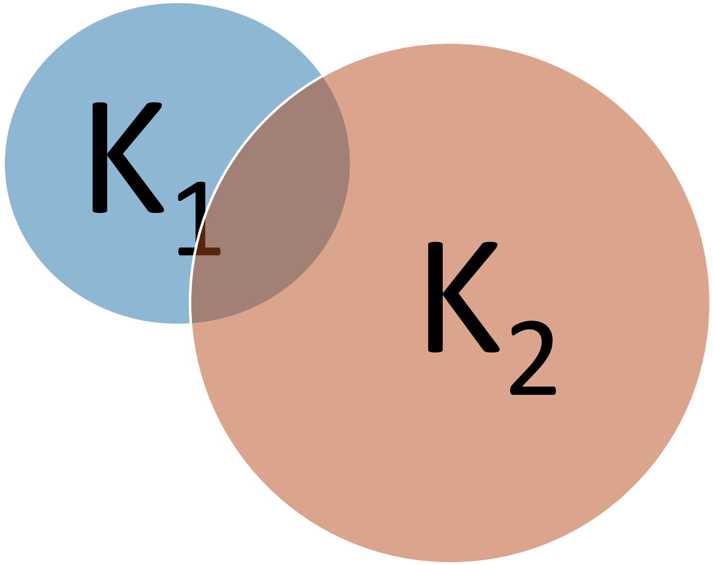
Conditional versus unconditional probability

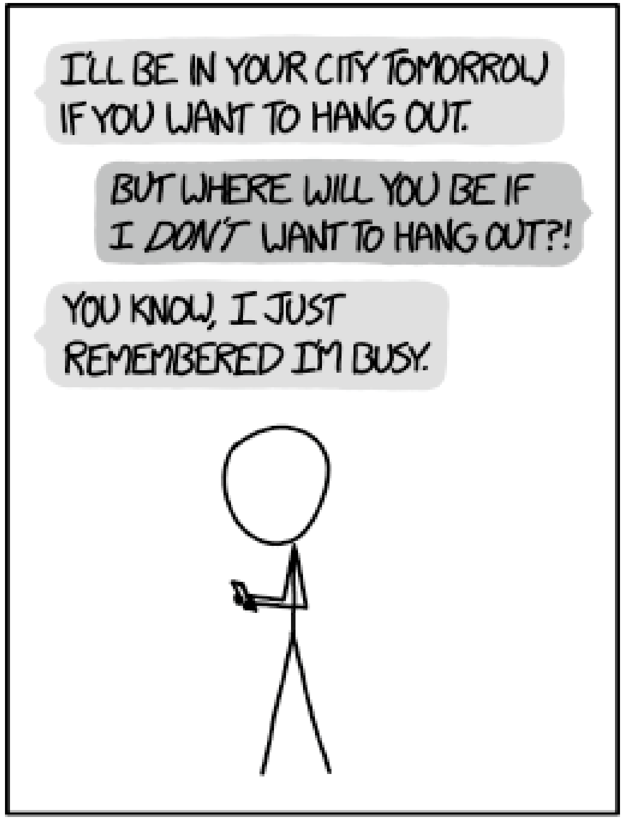
P(rain tomorrow)

P(rain tomorrow|it has rained today)

$P(\text{rain}|K_1)$

$P(\text{rain}|K_2)$





WHY I TRY NOT TO BE PEDANTIC ABOUT CONDITIONALS.

Independence

Event A is independent of E given K if

$P(A|K) = P(A|E \& K)$

or (dropping the condition on knowledge bases)

$P(A) = P(A|E)$

By symmetry

$P(E) = P(E|A)$

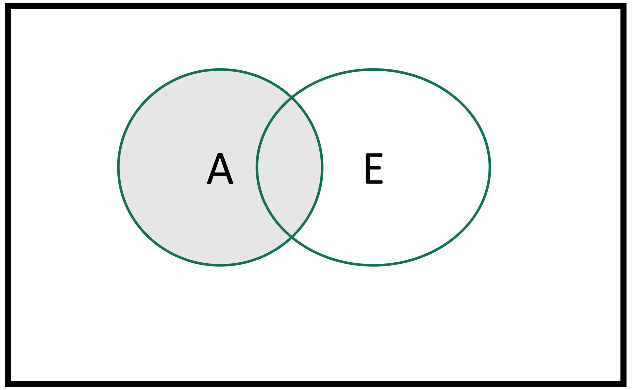
and

$P(A \& E) = P(A)P(E)$

More probability rules

Addition rule

P(A or E) = P(A) + P(E) – P(A & E)



Multiplication rule

P(A & E) = P(A|E)P(E)

P(A|E) -------------P(E|A)

Transposed conditionals or inverse conditional probability

Learning is an inverse problem

How to go from one to the other?

Learning

What we believe to begin with:

$P(A|K)$
$P(A)$

How to learn?

What we believe after evidence E is acquired:

$P(A|E \& K)$
$P(A|E)$

Bayes rule:

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)}$$

$P(E|A)$ is how likely it is to get E if A is true

Cromwell’s rule

”Think it possible you may be mistaken”

Bayes rule:

$P(F|E) = P(E|F)P(F) / P(E)$

What happens if $P(F) = 0$

You should not have probability 1 (or 0) for any event, other than one demonstrated by logic!

i.e. $0 < P(E|K) < 1$

but $P(E|K) = 1$ if and only if K logically implies the truth of E

Back to the BNs and assessing risk and impacts

Assessing risk and impacts

BNs support models

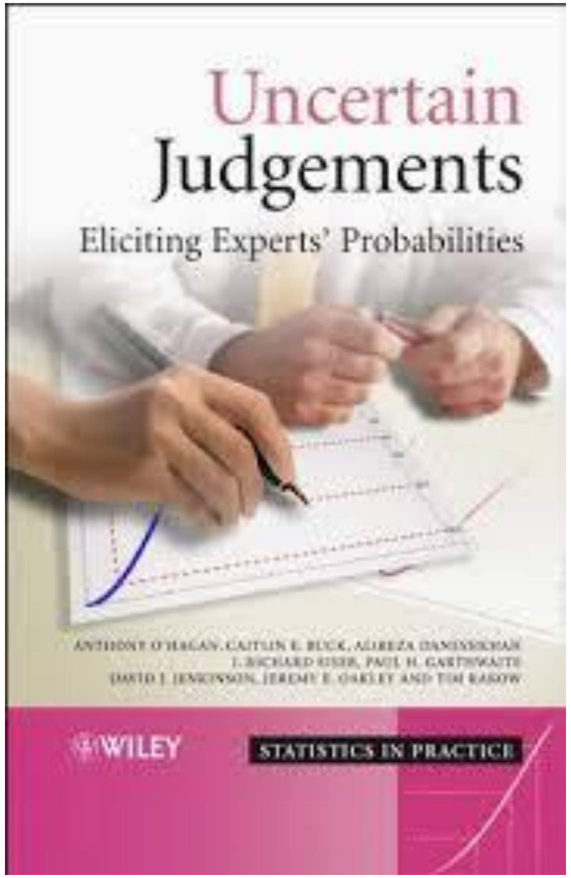
- Data-driven – learn parameters from data
- Expert-driven – parameters are assigned by experts
- Combinations of these (Hybrid approach)

The underlying theory

- Causal relationsships
- Mechansisms
- Processes - dynamics, time, space

- Directed Acyclic Graph
- Functions – determinstic links
- Assumptions

Expert knowledge – yes you are allowed to use that – but do it right!



An expert driven BN – classify invasiveness

A – Competitive ability in natural vegetation

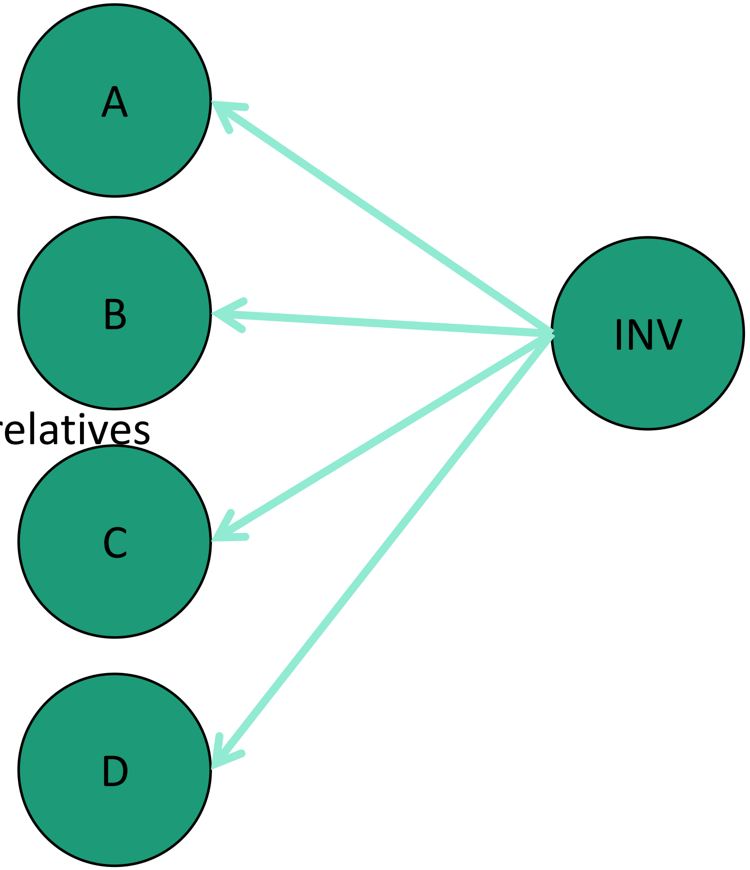
B – Population density

C – Realized dispersal capacity

D – Hybridization and gene flow to ”native” (old) relatives

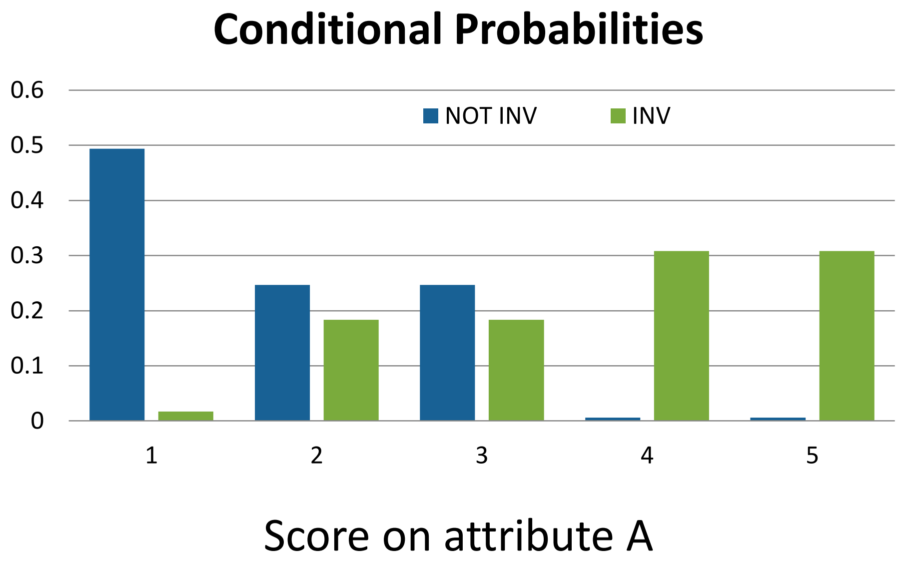
E – Time since introduction

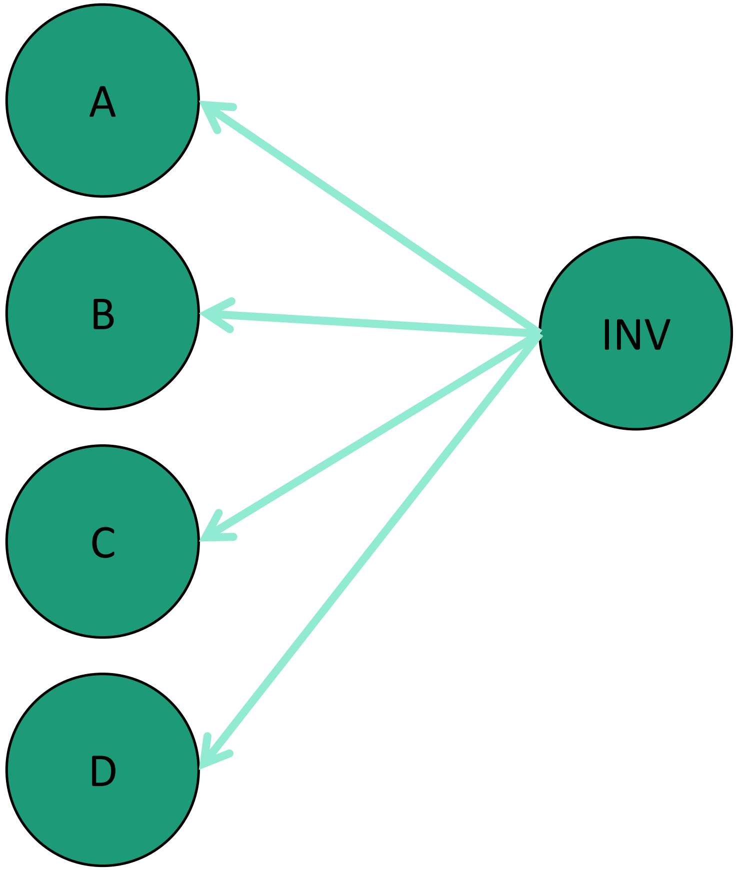
F – Distance to native range



LUND UNIVERSITY

An expert driven BN – classify invasiveness

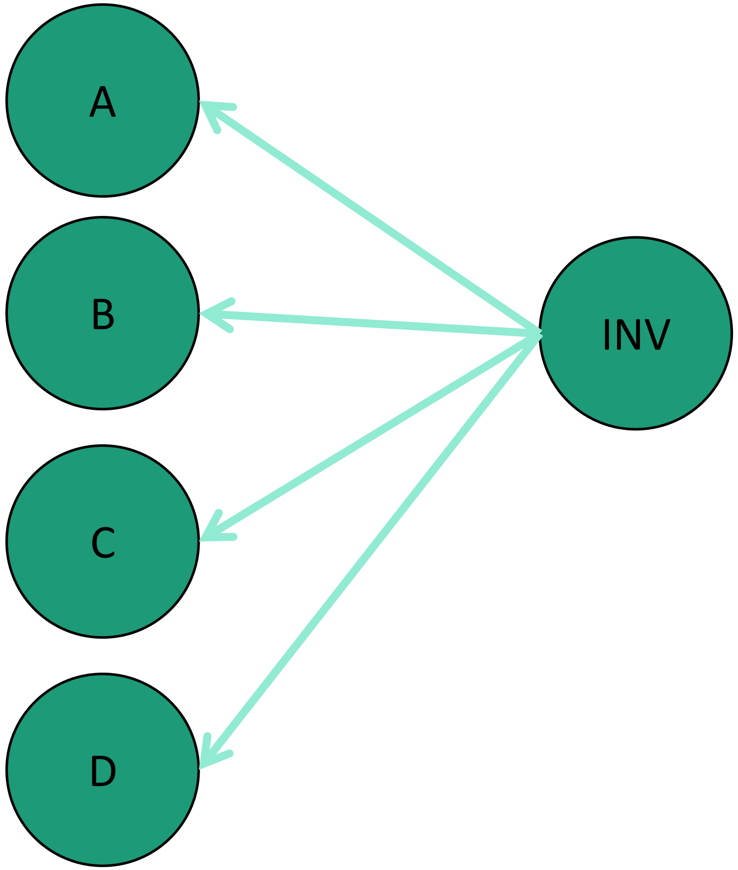




An expert driven BN – classify invasiveness

- What to do if the expert is uncertain about hes belief?
- What if different experts disagree?
- Is there a way to do this to avoid common errors?

- Structured approach to Expert’s Knowledge Elicitation
- Train the experts
- Use real experts



Make predictions for a few number of cases

- Importing case records from a data file cases…
- Mark all, Data -> Copy cases
- Open case manager
- Click on a case and Apply it
- Update the BN (F5)
- Study the bargraph of the Risk_Class Node

LUND UNIVERSITY

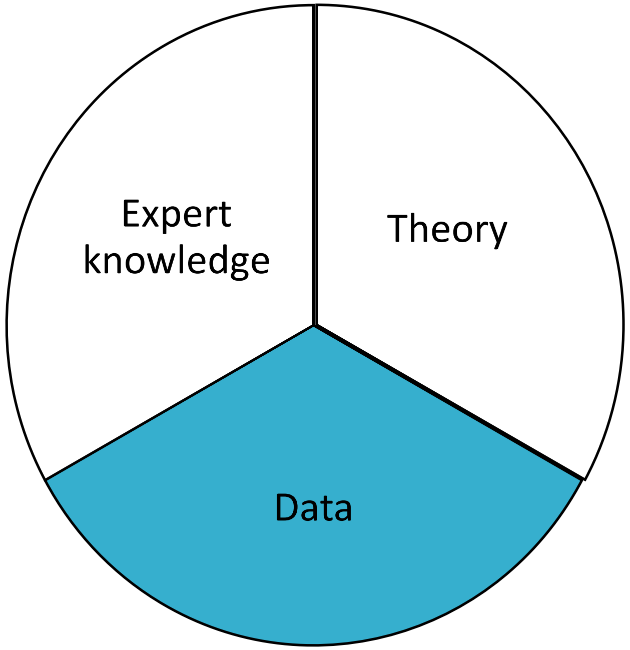
*A bit non-obvious way to do it when there is lots of cases

Make predictions*

- Open score file
- Data -> validate
- Set the Risk_class as fixed node
- Assign a file to save as output
- Open the output file and use the marginal probabilities

Lund University

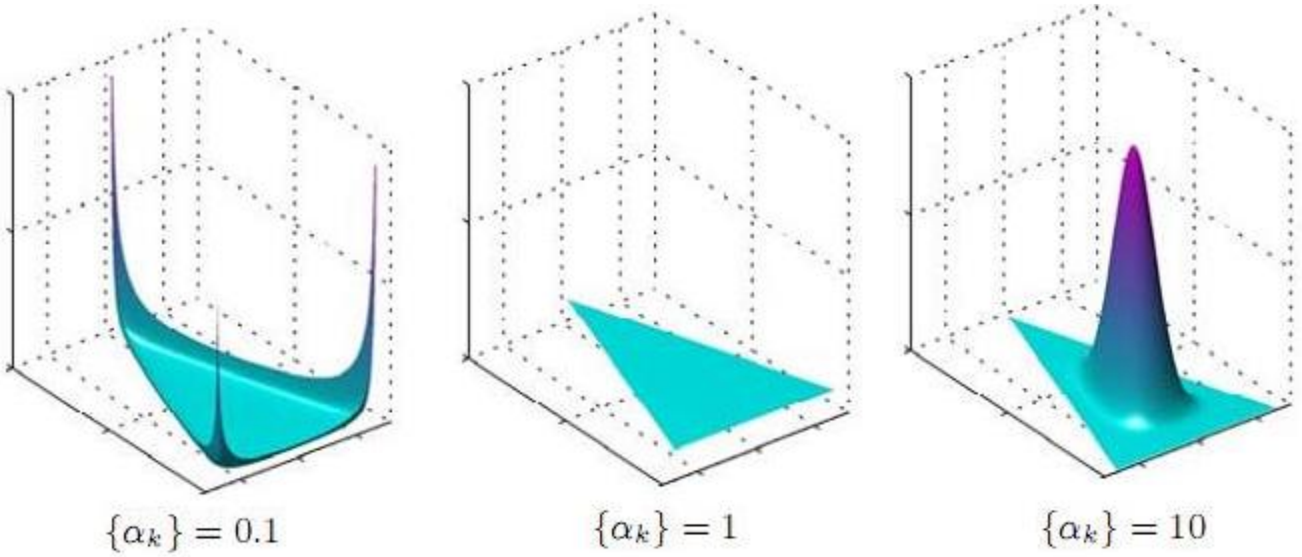
Learn from data – nice but beware of pitfalls



- Populate the network with probability distributions for parent nodes and conditional probability tables for child nodes using relative frequencies in data
- Weithing in prior probabilities
- Straightforward

Dirichlet distribution

- $Dir(\alpha)$
- Support:
- $x_j \in (0,1)$ and $\sum_{j=1}^{K} x_j = 1$
- Parameters:
- $\alpha = \{\alpha_1, \dots, \alpha_K\}$ and $\alpha_j > 0$
- Mean:
- $E(X_j) = \frac{\alpha_j}{\sum_1^K \alpha_k}$
- Marginal distribution:
- Beta distribution $X_j \sim Beta(\alpha_j, \left(\frac{\alpha_j}{\sum_1^K \alpha_k}\right) - \alpha_j)$
- Sometimes one use $s \cdot \{t_1, \dots, t_K\}$ instead of α



Useful finding

- ”The Dirichlet distribution form a conjugate family under multinomial sampling”
- Using Dir as prior for our belief in θ the posterior distribution is also Dir and easy to calculate
- This is true when the data is the result of *n* independent trials in which each trial result in one out of a fixed number of outcomes, e.g. outcome j occurs n_j times

$$s \to n + s$$

$$t_j \to \frac{n_j + st_j}{n + s}$$

Learning parameters* from data

- Load data file with configurations
- Learn parameters (set confidence weight on data)

*It is also possible to learn structure from data

Learning parameters from data

- Open training data file cancer_trainingdata.txt
- Select Data - > learn parameters
- Set confidence to 0 and tick uniformize (priors are uniform but learning only considers data)
- Set confidence to 1 and tick uniformize (priors are uniform and has low weight compared to data when learning)
- Repeat with cancer_trainingdata_small.txt

LUND UNIVERSITY

Confusion matrix

	Predicted condition	
True condition	Cancer	Not cancer
Cancer	TP	FN type II error
Not cancer	FP type I error	TN

Validate a BN

- Load data file with configurations
- Validate
  - Test only
  - Leave-one-out
  - K-fold cross validation
- Validation metrics
  - Accuracy
  - Confusion matrix
  - ROC curve