Introduction to Probabilistic Models

Silja Renooij & Antonio Salmerón Cerdán

Dept. of Information & Computing Sciences
Utrecht University
The Netherlands
s.renooij@uu.nl

Department of Mathematics
University of Almería
Spain
antonio.salmeron@ual.es

Trondheim, June 16 2025

Motivation

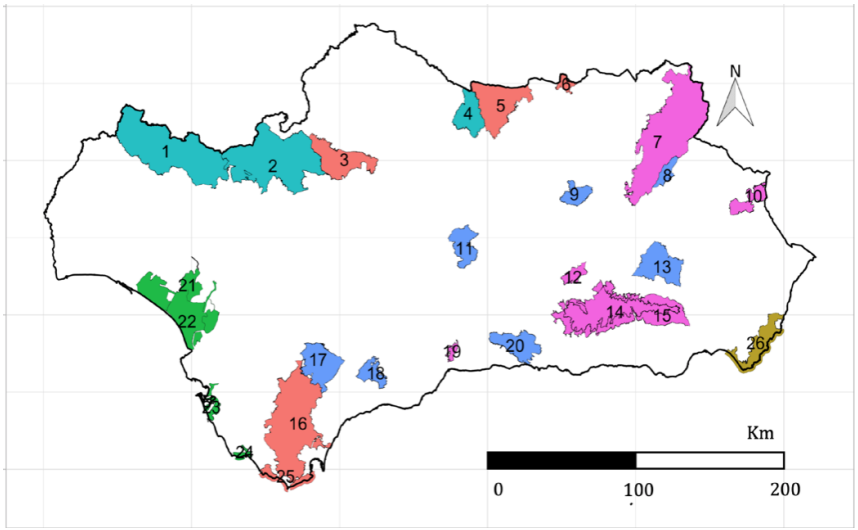
I always wondered how it would be if a Superior species landed on Earth and showed us how they played chess.
Now I know it.

Peter Heine Nielsen
Chess Grand Master and Magnus Carlsen's coach

The question is,

- Can we (humans) learn (interpret) anything from it?

Examples



Monitoring protected areas

Matching victims' DNA to relatives

Predicting events to avoid

Gen AI

Probabilistic models

All the previous examples:

- Operate in environments where large amounts of data are available
- However, data don't cover all the possible scenarios $\Rightarrow$ **UNCERTAINTY**
- Use a probabilistic model, typically learnt from data
- Use inference algorithms to carry out prediction and structure analysis

Probabilistic models offer at least:

- Principled quantification of uncertainty
- Natural way of dealing with missing data
- Interpretability

What we also need from them:

- Ability to operate in high dimensional spaces
- Support efficient inference and learning

Uncertainty

We often distinguish between two types of uncertainty:

- Epistemic: Due to lack of knowledge
- Aleatoric: Due to (pure) randomness, i.e. the variability in the outcome of an experiment due to random effects

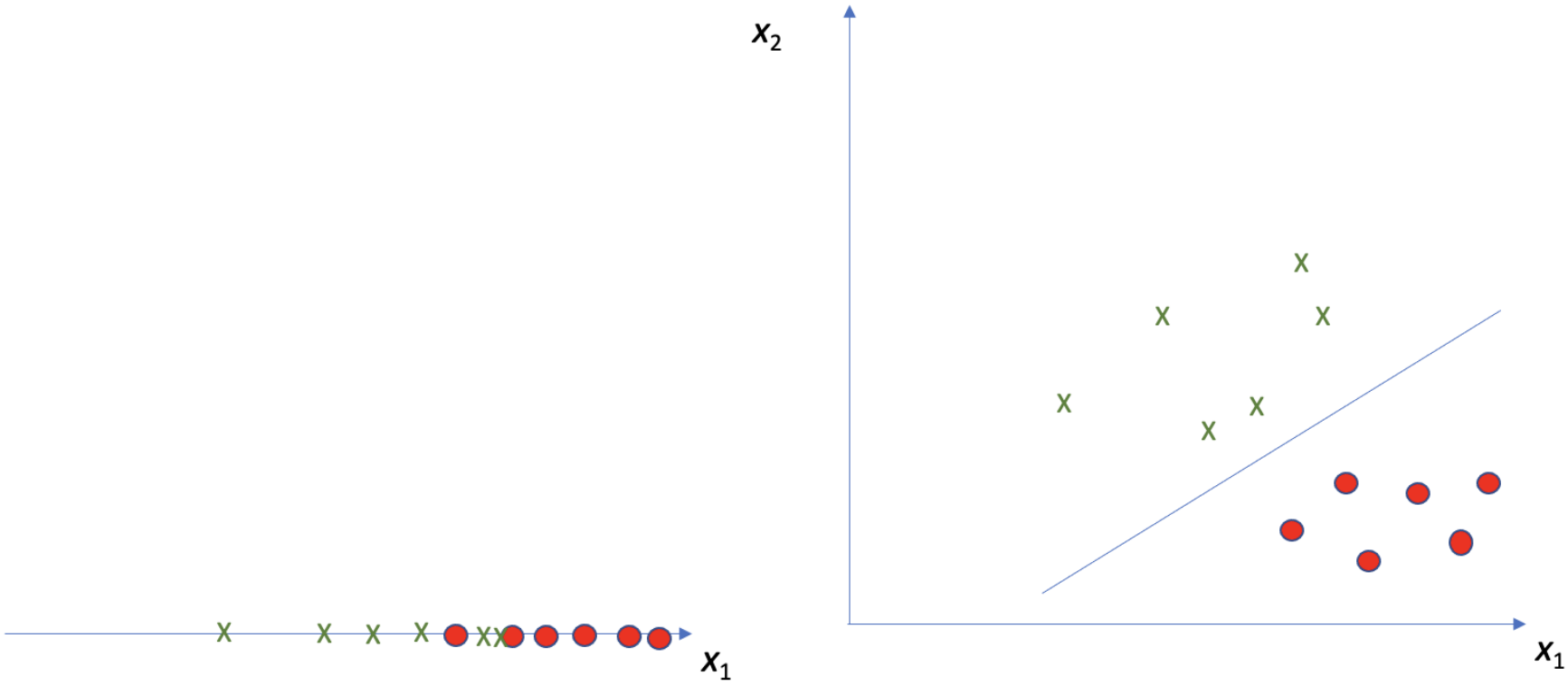
Example

- Assume we want to predict Y from X
- We estimate a joint distribution $p(x, y)$ [EPISTEMIC][REDUCIBLE]
- We predict Y using $p(y|x) = p(x, y)/p(x)$
- If we observe $X = x$, what does our model predict for Y? [ALEATORIC][IRREDUCIBLE]

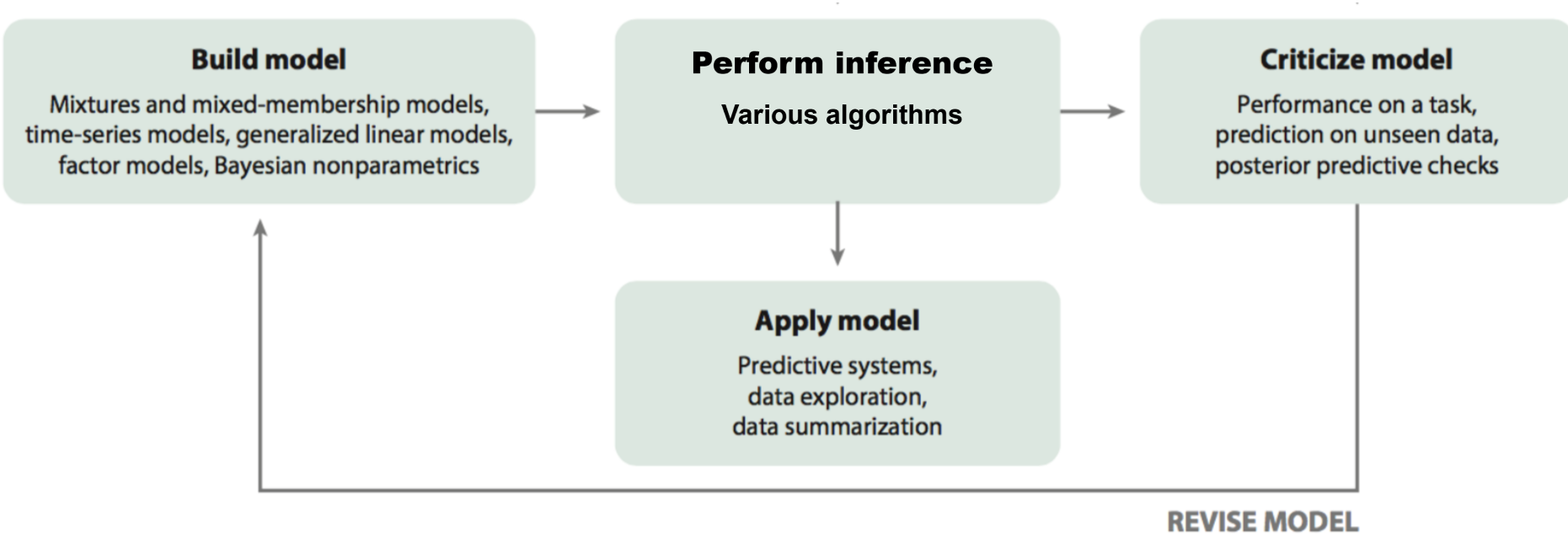
Uncertainty

Epistemic uncertainty can be reduced by

- gathering more data, but also by increasing the number of features



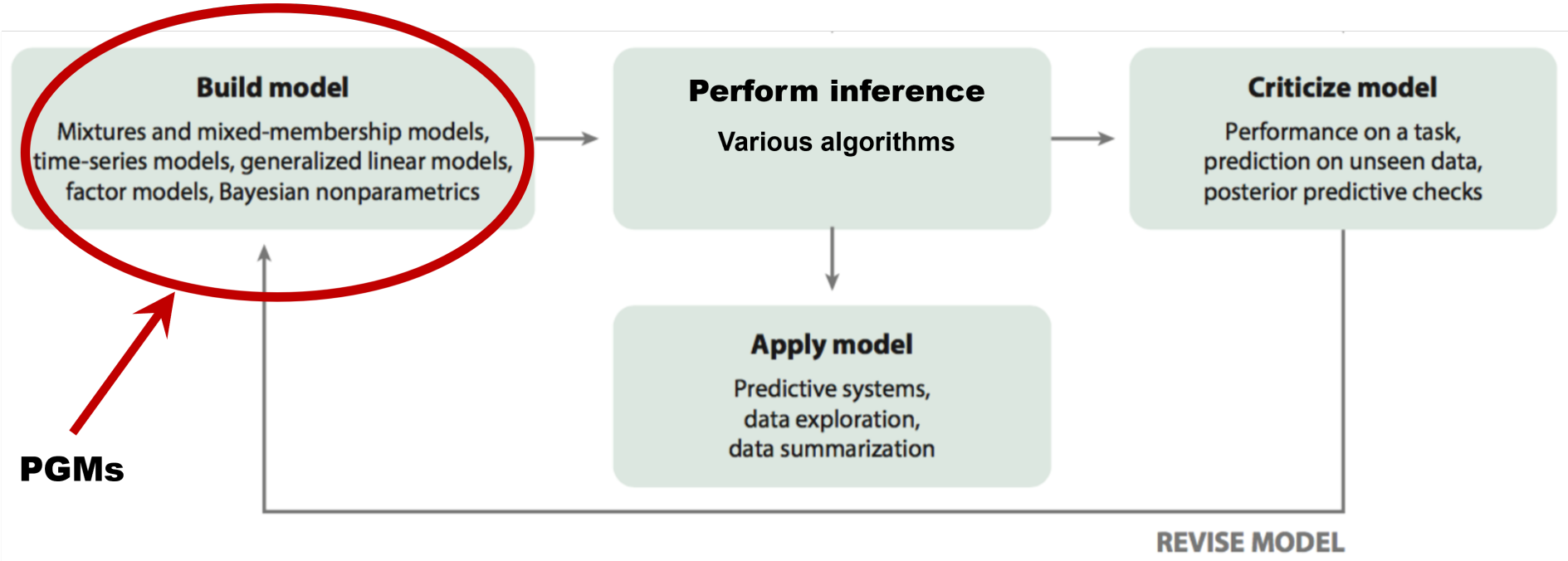
The Probabilistic Modelling Cycle



Trade-offs (performance, richness, costs); be clear about assumptions!

Adapted image from: David M. Blei (2014) "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and its Applications 1, 303–323.*

The Probabilistic Modelling Cycle



Probabilistic graphical models offer:

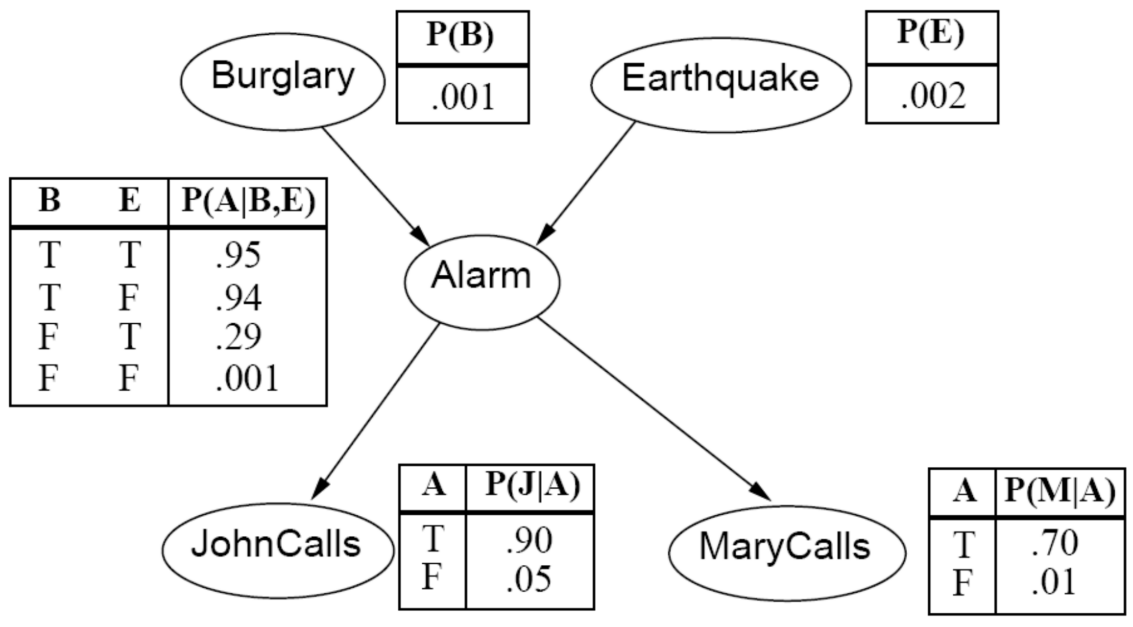
- Structured specification of high dimensional distributions in terms of low dimensional factors
- Efficient inference and learning, taking advantage of the structure
- Graphical representation interpretable by humans

Adapted image from: David M. Blei (2014) "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and its Applications 1, 303–323.*

Bayesian network: definition

A Bayesian network over random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ consists of

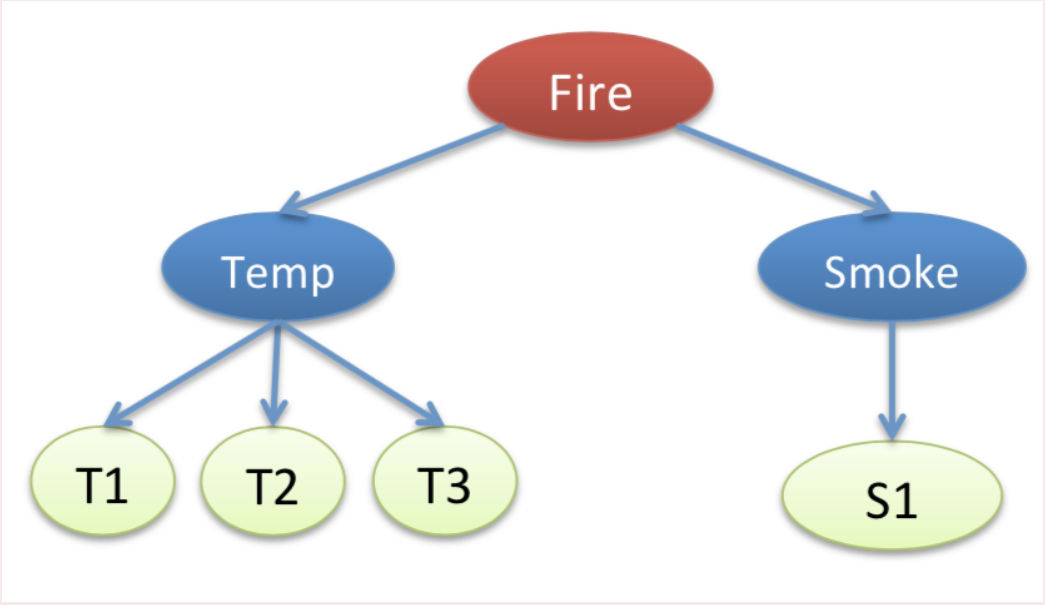
- A DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \boldsymbol{X}$
- A set of local conditional distributions $\mathcal{P} = \{\, p(X_i \mid pa(X_i)) \mid X_i \in \boldsymbol{X}\}$ where $pa(X_i)$ denotes the parents of X_i according to $\mathcal{E}$



Bayesian network: compact represention of the joint

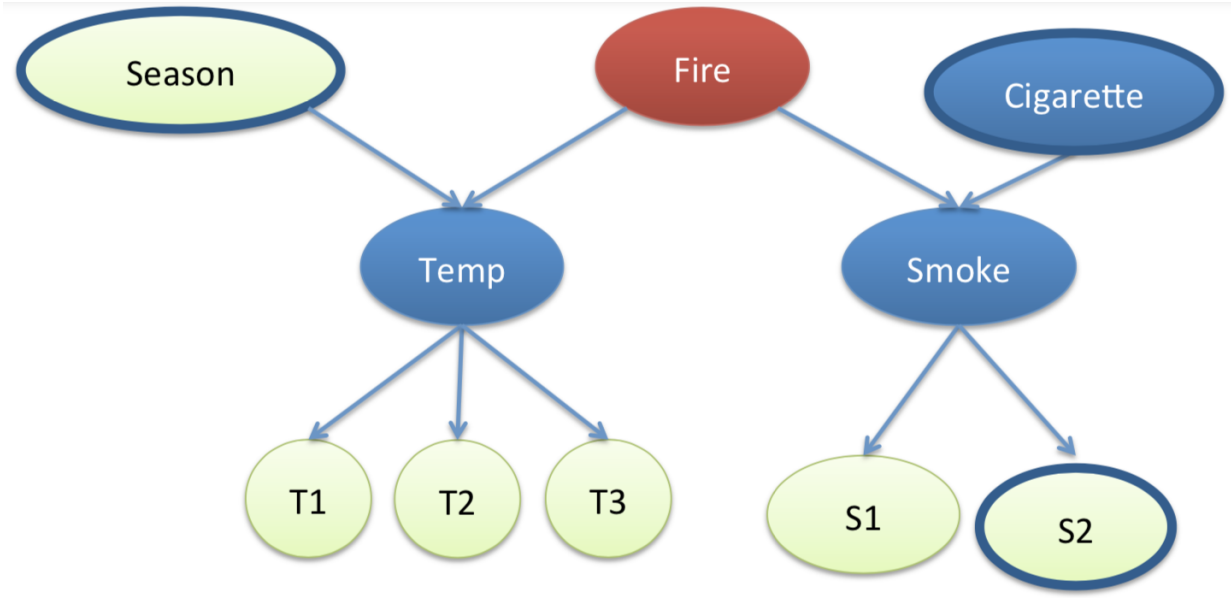
The DAG encodes independences among the variables (through d-separation); every Bayesian network therefore represents a factorized joint distribution:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid pa(X_i))$$



$$p(f, t, s, t_1, t_2, t_3, s_1) = p(t_1|t)p(t_2|t)p(t_3|t)p(s_1|s)p(t|f)p(s|f)p(f)$$

Bayesian networks: modular structure



$$p(se, f, c, t, s, t_1, t_2, t_3, s_1) = p(t_1|t)p(t_2|t)p(t_3|t)p(s_1|s)p(s_2|s)$$
$$p(t|se, f)p(s|f, c)p(se)p(f)p(c)$$

Inference in Bayesian networks

Assume a Bayesian network over variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$

From the joint distribution $p(X_1, \ldots, X_n)$ we can infer a.o.

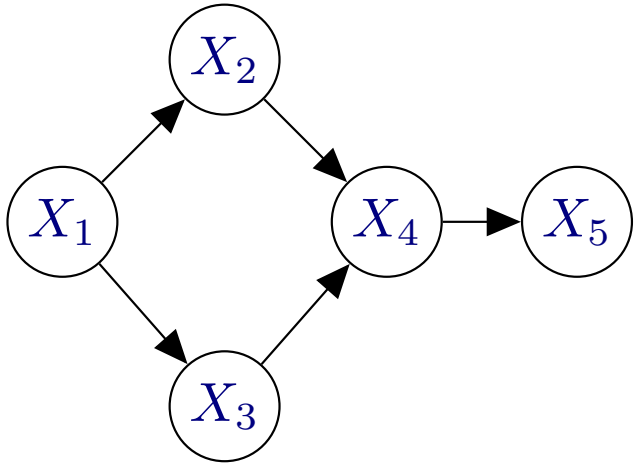
- the prior distribution $p(X_i)$ of any X_i,
- the posterior distribution $p(X_i \mid \boldsymbol{x}_E)$ of any X_i given evidence for $\boldsymbol{x}_E$.

Note: interpretation of these terms will differ when we consider learning!

Inference methods

- Exact
  - Brute force: compute $P(\boldsymbol{X}, \boldsymbol{x}_E)$ and marginalize out $\boldsymbol{X} \setminus \boldsymbol{X}_I$
  - Take advantage of the network structure
- Approximate
  - Sampling
  - Deterministic

Exact inference: Variable elimination



- We are interested in X_5
- All variables are discrete
- $E = \emptyset$

$$
\begin{aligned}
p(x_5) &= \sum_{x_1,\ldots,x_4} p(x_1, x_2, x_3, x_4, x_5) \\
&= \sum_{x_1,\ldots,x_4} p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_4) \\
&= \sum_{x_2,\ldots,x_4} \sum_{x_1} p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_4) \\
&= \sum_{x_2,\ldots,x_4} p(x_4|x_2, x_3)p(x_5|x_4) \boxed{\sum_{x_1} p(x_1)p(x_2|x_1)p(x_3|x_1)} \\
&= \sum_{x_2,\ldots,x_4} p(x_4|x_2, x_3)p(x_5|x_4)h(x_2, x_3)
\end{aligned}
$$

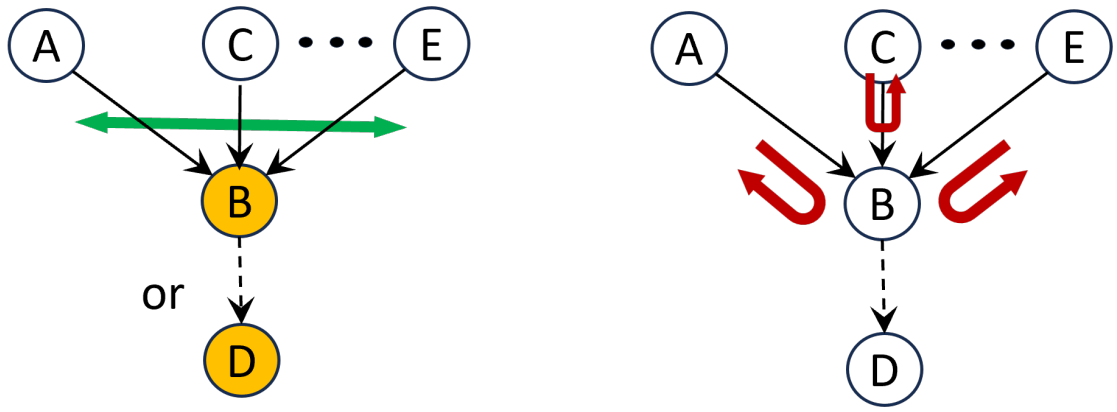
We have reached a similar problem as initially, but with one variable less.

Interpreting Bayesian network structures: d-separation

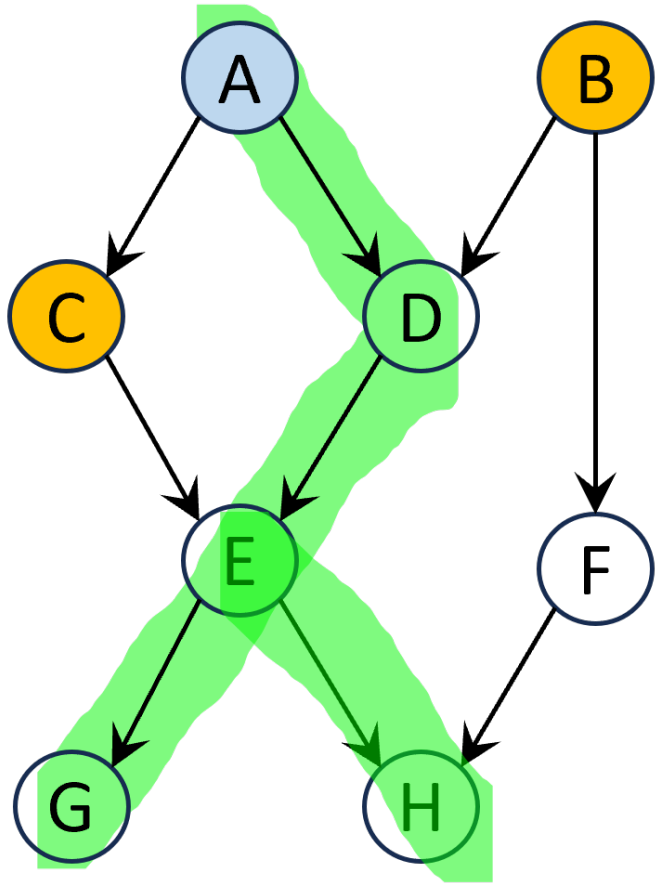
- Serial connection

- Diverging connection

- Converging connection

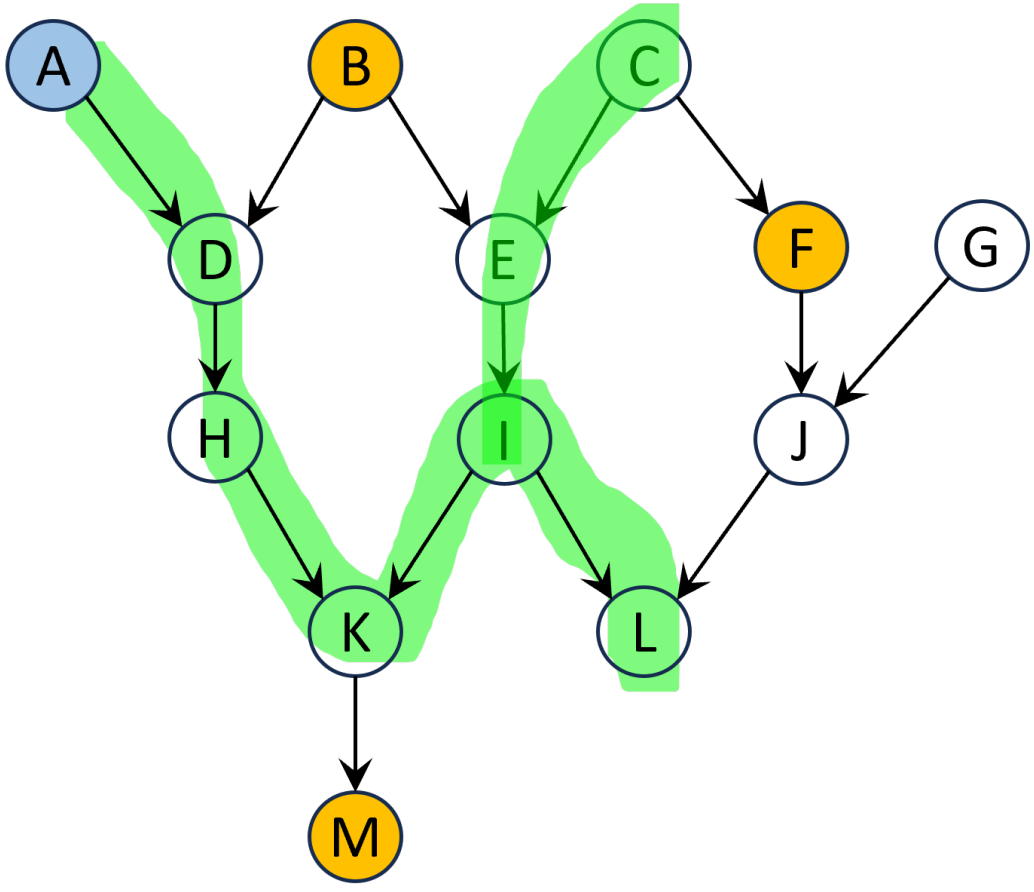


d-separation example - I



Which variables are d-separated from A given the evidence (in orange)?
All outside active (green) chains.

d-separation example - II



Which variables are d-separated from A given the evidence (in orange)?
All outside active (green) chains.

Monty Hall problem

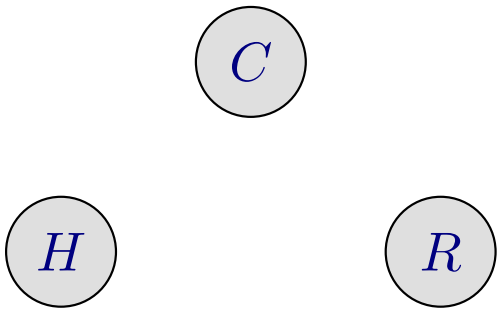
You are given the choice between 3 doors. One has a real prize behind it, the other two joke prizes.

You choose a door; the host then opens a door and offers you the choice to switch to a closed door.

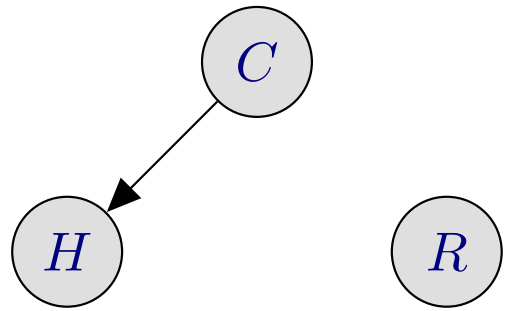
Would you switch?

Monty Hall problem: which DAG fits your assumptions?

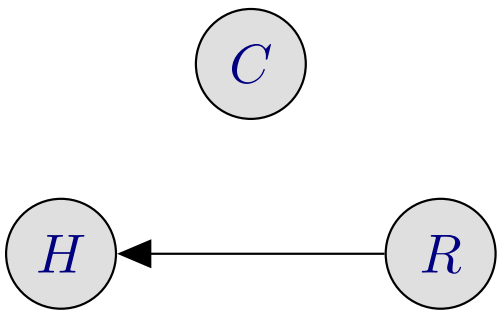
C: your choice of door; H: door opened by host Monty; R: door with real prize



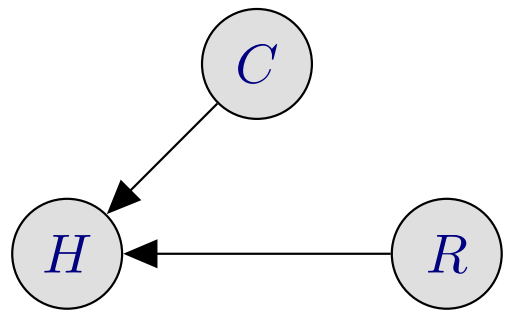
(I)



(II)

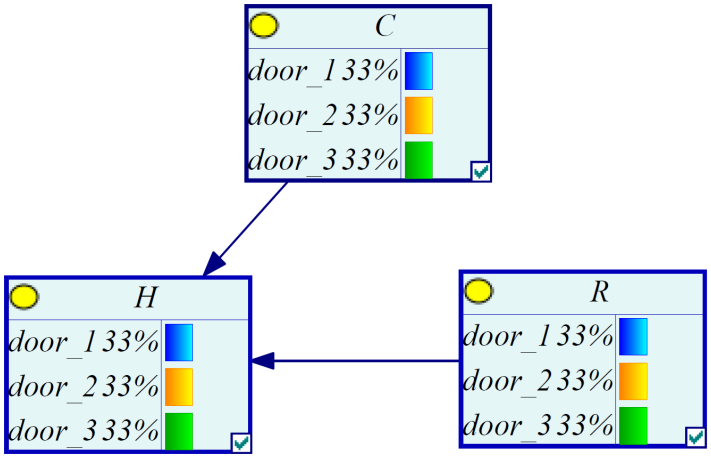


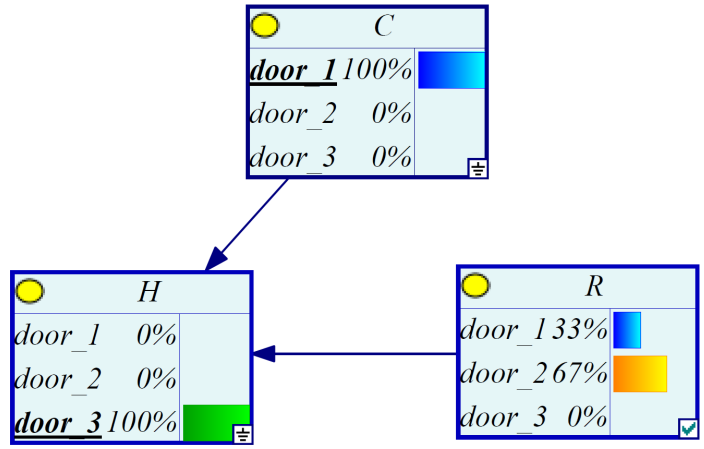
(III)



(IV)

Monty Hall: performing inference





$$p(H) = \sum_{c,r} p(H \mid c, r)p(c)p(r)$$

$$p(R \mid C = door_1, H = door_3) = \frac{p(H = door_3 \mid C = door_1, R)p(R)}{p(H = door_3)}$$

- What assumptions underlie your probability assessments?
- Get the DIY-MontyHall Jupyter notebook (e.g. through Google Colab) from https://github.com/probabilisticai/nordic-probai-2025

Probabilistic graphical models

Recall that probabilistic graphical models offer:

- Structured specification of high dimensional distributions in terms of low dimensional factors
- Graphical representation interpretable by humans
- Efficient inference and learning, taking advantage of the structure