Building Bayesian Networks
The focus today . . .

- Problem solving by Bayesian networks

- Designing Bayesian networks
  - Qualitative part (structure)
  - Quantitative part (probability assessment)

- Simplified Bayesian networks
  - In structure: Naïve Bayes, Tree-Augmented Networks
  - In probability assessment: Parent divorcing, Causal Independence
Problem solving

Bayesian networks: a declarative (knowing what) knowledge-representation formalism, i.e.,:

- mathematical basis
- problem to be solved determined by (1) entered evidence $\mathcal{E}$ (including potential decisions); (2) given hypothesis $H : P(H \mid \mathcal{E})$

Examples:

- Description of population (or prior information)
- Classification and diagnosis: $D = \arg \max_H P(H \mid \mathcal{E})$ i.e. $D$ is the hypothesis with maximum $P(H \mid \mathcal{E})$
- Prediction
- Decision making based on what-if scenario’s
Prior information

- Gives description of the population on which the assessed probabilities are based, i.e., the original probabilities before new evidence is uncovered.

- Marginal probabilities $P(V)$ for every vertex $V$, e.g., $P(\text{WILSON'S DISEASE} = \text{yes})$. 

Diagnostic problem solving

- Gives **description of the subpopulation** of the original population or individual cases

- Marginal probabilities $P^* (V) = P(V \mid \mathcal{E})$ for every vertex $V$, e.g., $P(\text{WILSON'S DISEASE} = \text{yes} \mid \mathcal{E})$ for entered evidence $\mathcal{E}$ (red vertices, with probability for one value equal to 1)
Prediction of associated findings

- Gives description of the findings associated with a given class or category, such as Wilson’s disease.

- Marginal probabilities $P^*(V) = P(V \mid \mathcal{E})$ for every vertex $V$, e.g., $P(\text{Kayser-Fleischer Rings} = \text{yes} \mid \mathcal{E})$ with $\mathcal{E}$ evidence.
Design of Bayesian network

Current design principle: start modelling qualitatively (different from traditional knowledge-based systems)

- causal graph
- qualitative probabilistic network
- quantitative network
- Bayesian network

- refinement
- variables/relationships
- domains of variables/qualitative probabilistic information
- numerical assessment
- experts/dataset
- evaluation
**Terminology**

- **Parent** SARS of Child FEVER
- **SARS** is Ancestor of TEMP
- **DYSPNOEA** is Descendant of VisitToChina
- **Query node**, e.g., FEVER
- **Evidence**, e.g., VisitToChina and TEMP
- **Markov blanket**, e.g.,
  
  for SARS: \{VisitToChina, DYSPNOEA, FEVER, FLU\}
Identify factors that are relevant

Determine how those factors are causally related to each other

The arc $\text{cause} \rightarrow \text{effect}$ does mean that $\text{cause}$ is a factor involved in causing $\text{effect}$
An effect that has two or more ingoing arcs from other vertices is a common effect of those causes.

Kinds of causal interaction:
- Synergy: POLUTION $\rightarrow$ CANCER $\leftarrow$ SMOKING
- Prevention: VACCINE $\rightarrow$ DEATH $\leftarrow$ SMALLPOX
- XOR: ALKALI $\rightarrow$ DEATH $\leftarrow$ ACID
A cause that has two or more outgoing arcs to other vertices is a common cause (factor) of those effects.

The effects of a common cause are usually observables (e.g. manifestations of failure of a device or symptoms in a disease).
FEVER and PNEUMONIA are two alternative causes of fever (but may enhance each other)

FLU has two common effects: MYALGIA and FEVER

High body TEMPerature is an indirect effect of FLU and PNEUMONIA, caused by FEVER
Check independence relationship

**Conditional independence:** $X \perp Y \mid Z$

- $\{\text{FEVER}\} \perp \{\text{MYALGIA}\} \mid \{\text{FLU}\}$
- $\odot \mid \{\text{FEVER}\}$
- $\{\text{PNEUMONIA}\} \perp \{\text{FLU}\} \mid \odot$
- $\{\text{PNEUMONIA}\} \not\perp \{\text{FLU}\} \mid \{\text{FEVER}\}$

![Graph with nodes FLU, MYALGIA, PNEUMONIA, FEVER, TEMP connected through arrows showing independence relationships.]
Choose variables

- Factors are **mutually exclusive** (cannot occur together with absolute certainty): put as values in the same variable, or

- Factors may co-occur: multiple variables

(a) Single variable

(b) Multiple variables
Choose values

- Discrete values
  - Mutually exclusive and exhaustive
  - Types:
    - binary, e.g., $\text{FLU} = \text{yes/no, true/false, 0/1}$
    - ordinal, e.g., $\text{INCOME} = \text{low, medium, high}$
    - nominal, e.g., $\text{COLOR} = \text{brown, green, red}$
    - integral, e.g., $\text{AGE} = \{1, \ldots, 120\}$
- Continuous values
- Discretization (of continuous values)
  - Example for $\text{TEMP}$:
    - $[-50, +5) \rightarrow \text{cold}$
    - $[+5, +20) \rightarrow \text{mild}$
    - $[+20, +50] \rightarrow \text{hot}$
Probability assessment

- **Qualitative assessment**
  - Check consistency
  - Indegree of vertex
    - ≤ 4
    - Quantitative assessment
      - Experts/dataset
      - Probability distribution
    - > 4
      - Divorcing/causal independence

- **Quantitative assessment**
  - Compare prediction with literature
  - Compare with test dataset

**Evaluation**
Expert judgements

- Qualitative probabilities:

- Qualitative orders:

<table>
<thead>
<tr>
<th>AGE</th>
<th>( P(\text{General Health Status} \mid \text{AGE}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-69</td>
<td>good &gt; average &gt; poor</td>
</tr>
<tr>
<td>70-79</td>
<td>average &gt; good &gt; poor</td>
</tr>
<tr>
<td>80-89</td>
<td>average &gt; poor &gt; good</td>
</tr>
<tr>
<td>( \geq 90 )</td>
<td>poor &gt; average &gt; good</td>
</tr>
</tbody>
</table>

- Equalities:

\[
P(\text{CANCER} = \text{T1} \mid \text{AGE} = 15 - 29) = \]
\[
P(\text{CANCER} = \text{T2} \mid \text{AGE} = 15 - 29) \]
Expert judgements (cont.)

Quantitative, subjective probabilities:

<table>
<thead>
<tr>
<th>AGE</th>
<th>good</th>
<th>average</th>
<th>poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-69</td>
<td>0.99</td>
<td>0.008</td>
<td>0.002</td>
</tr>
<tr>
<td>70-79</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>80-89</td>
<td>0.1</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>≥ 90</td>
<td>0.1</td>
<td>0.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>
A bottleneck in Bayesian networks

The number of parameters for the effect given \( n \) causes grows exponentially: \( \geq 2^n \) (for binary causes)

Unlikely evidence combination:

\[
P(fever | flu, rabies, ear_infection) = ?
\]

Problem: for many BNs too many probabilities have to be assessed
Solution: use simpler probabilistic model, such that either

- the structure becomes simpler, e.g.,
  - *naive* (independent) form BN
  - *Tree-Augmented Bayesian Network* (TAN)

or,

- the assessment of the conditional probabilities becomes simpler (even though the structure is still complex), e.g.,
  - *parent divorcing*
  - *causal independence* BN
Independent (Naive) form BN

- \( C \) is a class variable
- \( E_i \) are evidence variables and \( \mathcal{E} \subseteq \{E_1, \ldots, E_m\} \). We have \( E_i \perp \perp E_j \mid C \), for \( i \neq j \). Hence, using Bayes’ rule:

\[
P(C \mid \mathcal{E}) = \frac{P(\mathcal{E} \mid C)P(C)}{P(\mathcal{E})}
\]

with:

\[
P(\mathcal{E} \mid C) = \prod_{E \in \mathcal{E}} P(E \mid C)
\]

by cond. ind.

\[
P(\mathcal{E}) = \sum_{C} P(\mathcal{E} \mid C)P(C)
\]

marg. & cond.
### Example of Naive Bayes

**PlayTennis: training examples**

<table>
<thead>
<tr>
<th>Day</th>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Wind</th>
<th>PlayTennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D2</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D3</td>
<td>Overcast</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D4</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D5</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D6</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D7</td>
<td>Overcast</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D8</td>
<td>Sunny</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D9</td>
<td>Sunny</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D10</td>
<td>Rain</td>
<td>Mild</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D11</td>
<td>Sunny</td>
<td>Mild</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D12</td>
<td>Overcast</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D13</td>
<td>Overcast</td>
<td>Hot</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D14</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
</tbody>
</table>
### Example of Naive Bayes (1)

#### Learning phase

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>2/9</td>
<td>3/5</td>
</tr>
<tr>
<td>Overcast</td>
<td>4/9</td>
<td>0/5</td>
</tr>
<tr>
<td>Rain</td>
<td>3/9</td>
<td>2/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot</td>
<td>2/9</td>
<td>2/5</td>
</tr>
<tr>
<td>Mild</td>
<td>4/9</td>
<td>2/5</td>
</tr>
<tr>
<td>Cool</td>
<td>3/9</td>
<td>1/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Humidity</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3/9</td>
<td>4/5</td>
</tr>
<tr>
<td>Normal</td>
<td>6/9</td>
<td>1/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wind</th>
<th>Play=Yes</th>
<th>Play=No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>3/9</td>
<td>3/5</td>
</tr>
<tr>
<td>Weak</td>
<td>6/9</td>
<td>2/5</td>
</tr>
</tbody>
</table>

\[
P(\text{Play}=\text{Yes}) = \frac{9}{14} \quad P(\text{Play}=\text{No}) = \frac{5}{14}
\]
Example of Naive Bayes (2)

Testing phase (inference)

Evidence:
\[ x = (\text{Outlook}=\text{Sunny}, \text{Temperature}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=\text{Strong}) \]

Then given \( x \), PlayTennis=?
Example of Naive Bayes (3)

Testing phase (inference)

\[ P(Yes \mid x) = P(x \mid PlayTennis=Yes) \times P(PlayTennis=Yes) / P(x) \propto \]
\[ \propto [P(Sunny \mid Yes)P(Cool \mid Yes)P(High \mid Yes)P(Strong \mid Yes)]P(Play=Yes) = \]
\[ = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053 \]
\[ P(No \mid x) \propto [P(Sunny \mid No)P(Cool \mid No)P(High \mid No)P(Strong \mid No)]P(Play=No) = 0.0206 \]

Given that \( P(Yes \mid x') < P(No \mid x') \), then for \( x \) we label \text{PlayTennis} = \text{No}.

\textbf{Note:} to get probabilities we need to normalise

\[ P(Yes \mid x) = P(Yes \mid x) / (P(Yes \mid x) + P(No \mid x)) = 0.0053 / (0.0053 + 0.0206) = 0.20 \]
Tree-Augmented BN (TAN)

- Extension of Naive Bayes: reduce the number of independent assumptions
- Each node has at most two parents (one is the class node)
Divorcing multiple parents

(a) Original network

- Surgery
- Drug Treatment
- General Health
- Survival

(b) Divorced network

- Surgery
- Post-therapy Survival
- Drug Treatment
- General Health
- Survival

Reduction in number of probabilities to assess:

- Identify a potential common effect of two or more parent vertices of a vertex
- Introduce a new variable into the network, representing the common effect
Causal Independence

with:

- **cause variables** \( C_j \), **intermediate variables** \( I_j \), and the **effect variable** \( E \)

\[
P(E \mid I_1, \ldots, I_n) \in \{0, 1\}
\]

- **interaction function** \( f \), defined such that

\[
f(I_1, \ldots, I_n) = \begin{cases} 
e & \text{if } P(e \mid I_1, \ldots, I_n) = 1 \\ \neg e & \text{if } P(e \mid I_1, \ldots, I_n) = 0 \end{cases}
\]
Causal Independence: BN

\[ P(e \mid C_1, \ldots, C_n) = \sum_{I_1, \ldots, I_n} P(e \mid I_1, \ldots, I_n)P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) \]
\[ = \sum_{f(I_1, \ldots, I_n) = e} P(e \mid I_1, \ldots, I_n)P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) \]

Note that as \( I_i \perp \perp I_j \mid \emptyset \), and \( I_i \perp \perp C_j \mid C_i \), for \( i \neq j \), it holds that:

\[ P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) = \prod_{k=1}^{n} P(I_k \mid C_k) \]

Conclusion: assessment of \( P(I_i \mid C_i) \) instead of \( P(E \mid C_1, \ldots, C_n) \), i.e., \( 2n \) vs. \( 2^n \) probabilities
Causal independence: Noisy OR

Interactions among causes, as represented by the function $f$ and $P(E \mid I_1, I_2)$, is a logical OR.

Meaning: presence of any one of the causes $C_i$ with absolute certainty will cause the effect $e$ (i.e. $E = \text{true}$)

$$P(e \mid C_1, C_2) = ?$$
Causal independence: Noisy OR (cont.)

\[
P(e|C_1, C_2) = \sum_{I_1, I_2} P(e|I_1, I_2, C_1, C_2)P(I_1, I_2|C_1, C_2)
\]

= ?
Causal independence: Noisy OR (cont.)

\[
P(e|C_1, C_2) = \sum_{I_1, I_2} P(e|I_1, I_2, C_1, C_2)P(I_1, I_2|C_1, C_2)
\]

\[
= \sum_{f(I_1, I_2) = e} P(e|I_1, I_2) \prod_{k=1,2} P(I_k|C_k)
\]
Causal independence: Noisy OR (cont.)

\[ P(e|C_1, C_2) = \sum_{f(I_1, I_2) = e} P(e|I_1, I_2) \prod_{k=1,2} P(I_k|C_k) \]

\[ = P(i_1|C_1)P(i_2|C_2) + P(\neg i_1|C_1)P(i_2|C_2) + P(i_1|C_1)P(\neg i_2|C_2) \]
Noisy OR: Real-world example

Dynamic Bayesian network for predicting the development of hypertensive disorders during pregnancy

VascRisk (Vascular risk) has 11 causes and its original CPT requires the estimation of 20736 entries. Practically impossible!

Solution: use noisy OR to simplify it
Causal independence: Noisy AND

- Interactions among causes, as represented by the function $f$ and $P(E \mid I_1, I_2)$, is a logical AND.
- Meaning: presence of all causes $C_i$ with absolute certainty will cause the effect $e$ (i.e. $E = \text{true}$); otherwise, $\neg e$

$$P(e \mid C_1, C_2) = ?$$
Are Bayesian networks always suitable?

“Essentially, all models are wrong but some are useful”


- Problem (modelling) objective, e.g., for function approximation or pure numeric prediction without a need to explain the results a “black box” model such as neural networks can be sufficient
- Sufficient knowledge about the problem (domain experts, literature, data)
- Complexity of the problem e.g., is it decomposable
Refining causal graphs

Model refinement is necessary.

- **How:**
  - Manual
  - Automatic

- **What:**
  - Probability adjustment
  - Removing irrelevant factors
  - Adding previously hidden, unknown factors
  - Causal relationships adjustment, e.g., including, removing independence relations