Modeling the uncertainty.

Key challenges:

- How to represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - Humans can reason with uncertainty.

Pneumonia

? Paleness Fever Cough WBC count
Methods for representing uncertainty

**Probability theory**
- A well-defined theory for modeling and reasoning in the presence of uncertainty
- A natural choice to replace certainty factors

**Facts (propositional statements)**
- Are represented via random variables with two or more values
  
  **Example:** *Pneumonia* is a random variable
  
  **values:** *True* and *False*

- Each value can be achieved with some probability:
  
  \[ P(\text{Pneumonia} = \text{True}) = 0.001 \]
  
  \[ P(\text{WBC count} = \text{high}) = 0.005 \]

---

**Probability theory**

- Well-defined theory for representing and manipulating statements with uncertainty

**Axioms of probability:**

For any two propositions A, B.

1. \( 0 \leq P(A) \leq 1 \)
2. \( P(\text{True}) = 1 \) and \( P(\text{False}) = 0 \)
3. \( P(A \lor B) = P(A) + P(B) - P(A \land B) \)

---

![Diagram of sets A, A ∩ B, and B]
Methods for representing uncertainty

Probabilistic extension of propositional logic

• Propositions:
  – statements about the world
  – Statements are represented by the assignment of values to random variables

• Random variables:
  – Boolean
    - Pneumonia is either True, False
  – Multi-valued
    - Pain is one of \{Nopain, Mild, Moderate, Severe\}
  – Continuous
    - HeartRate is a value in <0;180>

Probabilities

Unconditional probabilities (prior probabilities)

\[ P(\text{Pneumonia}) = 0.001 \quad \text{or} \quad P(\text{Pneumonia} = \text{True}) = 0.001 \]
\[ P(\text{Pneumonia} = \text{False}) = 0.999 \]
\[ P(\text{WBC count} = \text{high}) = 0.005 \]

Probability distribution

• Defines probabilities for all possible value assignments to a random variable
• Values are mutually exclusive

\[
\begin{array}{c|c|c}
\text{Pneumonia} & \text{P}(\text{Pneumonia}) \\
\hline
\text{True} & 0.001 & 0.001 \\
\text{False} & 0.999 & 0.999 \\
\end{array}
\]
Probability distribution

Defines probability for all possible value assignments

**Example 1:**

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>P(Pneumonia)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.001</td>
</tr>
<tr>
<td>False</td>
<td>0.999</td>
</tr>
</tbody>
</table>

\[ P(\text{Pneumonia} = \text{True}) + P(\text{Pneumonia} = \text{False}) = 1 \]

Probabilities sum to 1 !!!

**Example 2:**

<table>
<thead>
<tr>
<th>WBCcount</th>
<th>P(WBCcount)</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>0.005</td>
</tr>
<tr>
<td>normal</td>
<td>0.993</td>
</tr>
<tr>
<td>low</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Joint probability distribution

Joint probability distribution (for a set variables)

- Defines probabilities for all possible assignments of values to variables in the set

**Example:** variables Pneumonia and WBCcount

\[ P(\text{pneumonia}, \text{WBCcount}) \]

Is represented by 2×3 array(matrix)

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>WBCcount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>True</td>
<td>0.0008</td>
</tr>
<tr>
<td>False</td>
<td>0.0042</td>
</tr>
</tbody>
</table>
Joint probability distribution

Joint probability distribution (for a set variables)
- Defines probabilities for all possible assignments of values to variables in the set

Example 2: Assume variables:
- Pneumonia (2 values)
- WBC count (3 values)
- Pain (4 values)

\[ P(\text{pneumonia}, \text{WBC count}, \text{Pain}) \]

is represented by \( 2 \times 3 \times 4 \) array

Example of an entry in the array
\[ P(\text{pneumonia} = T, \text{WBC count} = \text{high}, \text{Pain} = \text{severe}) \]

---

Joint probabilities: marginalization

Marginalization
- reduces the dimension of the joint distribution
- Sums variables out

\[ P(\text{pneumonia}, \text{WBC count}) \]

2 \times 3 matrix

\[
\begin{array}{ccc}
\text{WBC count} & \text{high} & \text{normal} & \text{low} \\
\text{Pneumonia} & & & \\
\text{True} & 0.0008 & 0.0001 & 0.0001 \\
\text{False} & 0.0042 & 0.9929 & 0.0019 \\
\hline
0.005 & 0.993 & 0.002 \\
\end{array}
\]

\[ P(\text{Pneumonia}) \]

\[ P(\text{WBC count}) \]

Marginalization (here summing of columns or rows)
Marginalization

Marginalization
• reduces the dimension of the joint distribution

\[ P(X_1, X_2, \ldots, X_{n-1}) = \sum_{\{X_n\}} P(X_1, X_2, \ldots, X_{n-1}, X_n) \]

• We can continue doing this

\[ P(X_1, \ldots, X_{n-2}) = \sum_{\{X_{n-1}, X_n\}} P(X_1, X_2, \ldots, X_{n-1}, X_n) \]

What is the maximal joint probability distribution?
• Full joint probability

Full joint distribution

• the joint distribution for all variables in the problem
  – It defines the complete probability model for the problem

Example: pneumonia diagnosis
• Variables: Pneumonia, Fever, Paleness, WBCcount, Cough
• Full joint probability: \( P(\text{Pneumonia}, \text{Fever}, \text{Paleness}, \text{WBCcount}, \text{Cough}) \)
  – defines the probability for all possible assignments of values to these variables

\[ P(\text{Pneumonia}=T, \text{WBCcount}=\text{High}, \text{Fever}=T, \text{Cough}=T, \text{Paleness}=T) \]
\[ P(\text{Pneumonia}=T, \text{WBCcount}=\text{High}, \text{Fever}=T, \text{Cough}=T, \text{Paleness}=F) \]
\[ P(\text{Pneumonia}=T, \text{WBCcount}=\text{High}, \text{Fever}=T, \text{Cough}=F, \text{Paleness}=T) \]
\[ \ldots \quad \text{etc} \]
• How many probabilities are there?
Full joint distribution

- **the joint distribution for all variables in the problem**
  - It defines the complete probability model for the problem

**Example:** pneumonia diagnosis

**Variables:** Pneumonia, Fever, Paleness, WBC count, Cough

Full joint probability: $P(\text{Pneumonia, Fever, Paleness, WBC count, Cough})$

- defines the probability for all possible assignments of values to these variables

$P(\text{Pneumonia}=T, \text{WBC count}= \text{High}, \text{Fever}=T, \text{Cough}=T, \text{Paleness}=T)$

$P(\text{Pneumonia}=T, \text{WBC count}= \text{High}, \text{Fever}=T, \text{Cough}=T, \text{Paleness}=F)$

$P(\text{Pneumonia}=T, \text{WBC count}= \text{High}, \text{Fever}=T, \text{Cough}=F, \text{Paleness}=T)$

... etc

- **How many probabilities are there?**
- **Exponential in the number of variables**

Full joint distribution

- **Any joint probability over a subset of variables can be obtained via marginalization**

$$P(\text{Pneumonia, WBC count, Fever}) = \sum_{c, p=(T,F)} P(\text{Pneumonia, WBC count, Fever, Cough}= c, \text{Paleness}= p)$$

- **Is it possible to recover the full joint from the joint probabilities over a subset of variables?**
Joint probabilities

- Is it possible to recover the full joint from the joint probabilities over a subset of variables?

\[ P(pneumonia, WBC count) \quad 2 \times 3 \text{ matrix} \]

<table>
<thead>
<tr>
<th>WBC count</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>False</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ P(WBC count) \]

\[ P(Pneumonia) \]

Joint probabilities and independence

- Is it possible to recover the full joint from the joint probabilities over a subset of variables?
- Only if the variables are independent !!!

\[ P(pneumonia, WBC count) \quad 2 \times 3 \text{ matrix} \]

<table>
<thead>
<tr>
<th>WBC count</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>False</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ P(WBC count) \]

\[ P(Pneumonia) \]
Variable independence

- The two events $A, B$ are said to be independent if:
  \[ P(A, B) = P(A)P(B) \]
- The variables $X, Y$ are said to be independent if their joint can be expressed as a product of marginals:
  \[ P(X, Y) = P(X)P(Y) \]

Conditional probabilities

- Conditional probability distribution
  \[ P(A \mid B) = ? \]
Conditional probabilities

- **Conditional probability distribution**
  \[ P(A \mid B) = \frac{P(A, B)}{P(B)} \text{ s.t. } P(B) \neq 0 \]

- **Product rule.** Join probability can be expressed in terms of conditional probabilities
  \[ P(A, B) = P(A \mid B)P(B) \]

- **Chain rule.** Any joint probability can be expressed as a product of conditionals
  \[ P(X_1, X_2, \ldots, X_n) = P(X_n \mid X_1, \ldots, X_{n-1})P(X_1, \ldots, X_{n-1}) \]
  \[ = P(X_n \mid X_1, \ldots, X_{n-1})P(X_{n-1} \mid X_1, \ldots, X_{n-2})P(X_1, \ldots, X_{n-2}) \]
  \[ = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \]
Conditional probabilities

**Conditional probability**
- Is defined in terms of the joint probability:
  \[ P(A \mid B) = \frac{P(A, B)}{P(B)} \text{ s.t. } P(B) \neq 0 \]
- **Example:**
  \[ P(\text{pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) = \frac{P(\text{pneumonia} = \text{true, WBC count} = \text{high})}{P(\text{WBC count} = \text{high})} \]
  \[ P(\text{pneumonia} = \text{false} \mid \text{WBC count} = \text{high}) = \frac{P(\text{pneumonia} = \text{false, WBC count} = \text{high})}{P(\text{WBC count} = \text{high})} \]

**Conditional probability distribution**
- Defines probabilities for all possible assignments, given a fixed assignment to some other variable values
  \[ P(\text{Pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) \]

\[
\begin{array}{c|cc|c}
\text{WBC count} & \text{True} & \text{False} & \text{Pneumonia} \\
\hline
\text{high} & 0.08 & 0.92 & 1.0 \\
\text{normal} & 0.0001 & 0.9999 & 1.0 \\
\text{low} & 0.0001 & 0.9999 & 1.0 \\
\end{array}
\]

Variable we condition on
\[ P(\text{Pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) \]
\[ + P(\text{Pneumonia} = \text{false} \mid \text{WBC count} = \text{high}) \]
Bayes rule

**Conditional probability.**

\[
P(A \mid B) = \frac{P(A, B)}{P(B)} \quad P(A, B) = P(B \mid A)P(A)
\]

**Bayes rule:**

\[
P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}
\]

**When is it useful?**

- When we are interested in computing the diagnostic query from the causal probability

\[
P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}
\]

- **Reason:** It is often easier to assess causal probability
  - E.g. Probability of pneumonia causing fever
  vs. probability of pneumonia given fever

---

**Bayes Rule in a simple diagnostic inference**

- **Device** (equipment) operating *normally* or *malfunctioning*.
  - Operation of the device sensed indirectly via a sensor
- **Sensor reading** is either *High* or *Low*

<table>
<thead>
<tr>
<th>Status \ Sensor</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>malfunctioning</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(Device status)</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
</tr>
<tr>
<td>malfunctioning</td>
</tr>
</tbody>
</table>
Bayes Rule in a simple diagnostic inference.

- **Diagnostic inference**: compute the probability of device operating normally or malfunctioning given a sensor reading

\[
P(\text{Device status} | \text{Sensor reading} = \text{high}) = ?
\]

\[
= \left( \frac{P(\text{Device status} = \text{normal} | \text{Sensor reading} = \text{high})}{P(\text{Device status} = \text{malfunctioning} | \text{Sensor reading} = \text{high})} \right)
\]

- Note that typically the opposite conditional probabilities are given to us: they are much easier to estimate
- **Solution**: apply Bayes rule to reverse the conditioning variables

---

Bayes Rule in a simple diagnostic inference

- **Device** (equipment) operating *normally or malfunctioning*.
  - Operation of the device sensed indirectly via a sensor
- **Sensor reading** is either *High* or *Low*

<table>
<thead>
<tr>
<th>Status</th>
<th>Sensor</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>malfunc</td>
<td>0.6</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

\[
P(\text{Device status} | \text{Sensor reading} = \text{high}) = ?
\]
Bayes rule

Assume a variable $A$ with multiple values $a_1, a_2, \ldots, a_k$

Bayes rule can be rewritten as:

$$P(A = a_j \mid B = b) = \frac{P(B = b \mid A = a_j)P(A = a_j)}{P(B = b)}$$

$$= \frac{P(B = b \mid A = a_j)P(A = a_j)}{\sum_{i=1}^{k} P(B = b \mid A = a_i)P(A = a_i)}$$

Used in practice when we want to compute:

$$P(A \mid B = b)$$

for all values of $a_1, a_2, \ldots, a_k$

---

Probabilistic inference

Various inference tasks:

- **Diagnostic task. (from effect to cause)**
  $$P(\text{Pneumonia} \mid \text{Fever} = T)$$

- **Prediction task. (from cause to effect)**
  $$P(\text{Fever} \mid \text{Pneumonia} = T)$$

- **Other probabilistic queries** (queries on joint distributions).
  $$P(\text{Fever})$$
  $$P(\text{Fever, ChestPain})$$
Inference

Any query can be computed from the full joint distribution !!!

- **Joint over a subset of variables** is obtained through marginalization

\[ P(A = a, C = c) = \sum_i \sum_j P(A = a, B = b_i, C = c, D = d_j) \]

- **Conditional probability over set of variables**, given other variables’ values is obtained through marginalization and definition of conditionals

\[
P(D = d \mid A = a, C = c) = \frac{P(A = a, C = c, D = d)}{P(A = a, C = c)} = \frac{\sum_i P(A = a, B = b_i, C = c, D = d)}{\sum_j \sum_i P(A = a, B = b_i, C = c, D = d_j)}
\]

Inference

Any query can be computed from the full joint distribution !!!

- Any joint probability can be expressed as a product of conditionals via the **chain rule**.

\[
P(X_1, X_2, \ldots X_n) = P(X_n \mid X_1, \ldots X_{n-1})P(X_1, \ldots X_{n-1})
\]

\[= P(X_n \mid X_1, \ldots X_{n-1})P(X_{n-1} \mid X_1, \ldots X_{n-2})P(X_1, \ldots X_{n-2}) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots X_{i-1})
\]

- Sometimes it is easier to define the distribution in terms of conditional probabilities:
  - E.g. \( P(\text{Fever} \mid \text{Pneumonia} = T) \)
  - \( P(\text{Fever} \mid \text{Pneumonia} = F) \)
Modeling uncertainty with probabilities

- Defining the **full joint distribution** makes it possible to represent and reason with uncertainty in a uniform way.
- We are able to handle an arbitrary inference problem.

**Problems:**

- **Space complexity.** To store a full joint distribution we need to remember $O(d^n)$ numbers.
  - $n$ – number of random variables, $d$ – number of values.
- **Inference (time) complexity.** To compute some queries requires $O(d^n)$ steps.
- **Acquisition problem.** Who is going to define all of the probability entries?

---

Medical diagnosis example

- **Space complexity.**
  - Number of assignments: $2 \times 2 \times 2 \times 3 \times 2 = 48$.
  - We need to define at least 47 probabilities.

- **Time complexity.**
  - Assume we need to compute the marginal of Pneumonia=T from the full joint.

\[
P(Pneumonia = T) = \sum_{i \in \{T,F\}} \sum_{j \in \{T,F\}} \sum_{k = h,n,l} \sum_{u \in \{T,F\}} P(Fever = i, Cough = j, WBCcount = k, Pale = u) \]

  - Sum over: $2 \times 2 \times 3 \times 2 = 24$ combinations.
Modeling uncertainty with probabilities

- **Knowledge based system era (70s – early 80’s)**
  - Extensional non-probabilistic models
  - Solve the space, time and acquisition bottlenecks in probability-based models
  - froze the development and advancement of KB systems and contributed to the slow-down of AI in 80s in general

- Breakthrough (late 80s, beginning of 90s)
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Partial solutions for time complexities
  - Bayesian belief network

---

Bayesian belief networks (BBNs)

**Bayesian belief networks.**
- Represent the full joint distribution over the variables more compactly with a **smaller number of parameters**.
- Take advantage of **conditional and marginal independences** among random variables

- **A and B are independent**
  
  \[ P(A, B) = P(A)P(B) \]

- **A and B are conditionally independent given C**
  \[ P(A, B | C) = P(A | C)P(B | C) \]
  \[ P(A | C, B) = P(A | C) \]