Uncertainty.
Bayesian belief networks.

Modeling the uncertainty.

Key challenges:
- How to represent nondeterministic (stochastic) relations?
- How to manipulate such knowledge to make inferences?
  - Humans can reason with uncertainty.
Modeling uncertainty with probabilities

Probabilistic extension of propositional logic.

- **Propositions:**
  - statements about the world
  - Represented by the assignment of values to random variables

- **Random variables:**
  - **Boolean**
    - Pneumonia is either True, False
      - Random variable: Pneumonia
      - Values: True, False
  - **Multi-valued**
    - Pain is one of {Nopain, Mild, Moderate, Severe}
      - Random variable: Pain
      - Values: Nopain, Mild, Moderate, Severe
  - **Continuous**
    - HeartRate is a value in <0; 250>
      - Random variable: HeartRate
      - Values: Continuous

---

Probabilities

Unconditional probabilities

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

Probability distribution

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

<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>
Probability distribution

Defines probability for all possible value assignments

**Example 1:**

\[
P(P\text{neumonia} = \text{True}) = 0.001 \\
P(P\text{neumonia} = \text{False}) = 0.999
\]

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

Probabilities sum to 1 !!!

**Example 2:**

\[
P(WBC\text{count} = \text{high}) = 0.005 \\
P(WBC\text{count} = \text{normal}) = 0.993 \\
P(WBC\text{count} = \text{high}) = 0.002
\]

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

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

Is represented by 2×3 matrix

<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>

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

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

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

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

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

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

\[
\text{Marginalization (here summing of columns or rows)}
\]

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

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

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 defines the probability for all possible assignments of values to Pneumonia, Fever, Paleness, WBCcount, Cough

5 variables: full joint is captured by a 5-dimensional table

Full joint distribution

• Any joint probability for a subset of variables can be obtained from the full joint via marginalization

\[
P(Pneumonia, WBCcount, Fever) = \sum_{c, p \in \{T, F\}} P(Pneumonia, WBCcount, Fever, Cough = c, Paleness = p)
\]

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

- Any joint probability for a subset of variables can be obtained from the full joint via marginalization

\[
P(P\text{neumonia}, WBC\text{count}, Fever) = \sum_{c, p \in \{T, F\}} P(P\text{neumonia}, WBC\text{count}, Fever, Cough = c, Paleness = p)
\]

- Is it possible to recover the joint distribution for a set of variables from joint probabilities defined for its subsets?

Relations among joint distributions

Assume:

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Pneumonia)</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P(Fever)</td>
<td>0.05</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Can we unambiguously compute the joint over the two variables?

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Fever, Pneumonia)</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Relations among joint distributions

Assume:

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\text{Pneumonia})$</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>$P(\text{Fever})$</td>
<td>0.05</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Can we unambiguously compute the joint over the two variables?

No! More than one probability value is possible for joint table entries.

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\text{Fever, Pneumonia})$</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Relations among joint distribution

Assume:

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\text{Pneumonia})$</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>$P(\text{Fever})$</td>
<td>0.05</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Can we unambiguously compute the joint over the two variables?

The joint has more free parameters, the two individual distributions together.

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\text{Fever, Pneumonia})$</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
### Relations among joint distribution

**Assume:**

<table>
<thead>
<tr>
<th>Event</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Pneumonia)</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>0.05</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Is there a condition that would let us unambiguously compute the joint over two variables?

Yes. When the two variables are independent!

<table>
<thead>
<tr>
<th>Event</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Fever, Pneumonia)</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Relations among joint distribution

Assume:

\[
\begin{array}{c|cc}
 & \text{True} & \text{False} \\
P(\text{Pneumonia}) & 0.001 & 0.999 \\
P(\text{Fever}) & 0.05 & 0.95 \\
\end{array}
\]

- 1 free parameter

\[
P(\text{Pneumonia=True}) \times P(\text{Fever=True})
\]

\[
\begin{array}{c|cc}
 & \text{True} & \text{False} \\
P(\text{Fever, Pneumonia}) & 0.00005 & 0.0495 \\
 & 0.0095 & ? \\
\end{array}
\]

- 3 free parameters

Conditional probabilities

**Conditional probability distribution**

- Defines probabilities for all possible assignments, given a fixed assignment to some other variable values

\[
P(\text{Pneumonia = true} | \text{WBCcount} = \text{high})
\]

\[
P(\text{Pneumonia} | \text{WBCcount}) \quad \text{3 element vector of 2 elements}
\]

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.08</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>False</td>
<td>0.92</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

\[
P(\text{Pneumonia = true} | \text{WBCcount} = \text{high}) + P(\text{Pneumonia = false} | \text{WBCcount} = \text{high})
\]
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:
\[
P(pneumonia= true \mid WBC\text{count}= high) = \frac{P(pneumonia= true, WBC\text{count}= high)}{P(WBC\text{count}= high)}
\]
\[
P(pneumonia= false \mid WBC\text{count}= high) = \frac{P(pneumonia= false, WBC\text{count}= high)}{P(WBC\text{count}= high)}
\]

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
\[
P(X_1, X_2, \ldots X_n) = P(X_n \mid X_1, \ldots X_{n-1})P(X_1, \ldots X_{n-1})
\]
\[
= P(X_n \mid X_1, \ldots X_{n-1})P(X_{n-1} \mid X_1, \ldots X_{n-2})P(X_1, \ldots X_{n-2})
\]
\[
= \prod_{i=1}^n P(X_i \mid X_1, \ldots X_{i-1})
\]
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>Device status</th>
<th>P(Device status)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>0.9</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>malfunctioning</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Device\Sensor</th>
<th>P(Sensor reading</th>
<th>Device status)</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>high</td>
<td>0.1</td>
</tr>
<tr>
<td>malfunctioning</td>
<td>low</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Device\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>
Bayes Rule in a simple diagnostic inference.

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

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

- 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

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

**Bayes rule can be rewritten as**:

\[
P(A = a_j | B = b) = \frac{P(B = b | A = a_j)P(A = a_j)}{P(B = b)}
\]

\[
= \frac{\sum_{i=1}^{k} P(B = b | A = a_j)P(A = a_j)}{\sum_{i=1}^{k} P(B = b | A = a_j)P(A = a_j)}
\]

Used in practice when we want to compute:

\[
P(A | B = b) \quad \text{for all values of } a_1, a_2, \ldots a_k
\]
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
  \[ P(\text{A} = a, \text{C} = c) = \sum_{i} \sum_{j} P(A = a, B = b_i, C = c, D = d_j) \]

• **Conditional probability over set of variables**, given other variables’ values is obtained through marginalization and definition of conditionals
  \[ P(D = d \mid A = a, C = c) = \frac{P(A = a, C = c, D = d)}{P(A = a, C = c)} = \frac{\sum P(A = a, B = b_i, C = c, D = d)}{\sum \sum P(A = a, B = b_i, C = c, D = d_j)} \]
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.

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

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

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

- Sometimes it is easier to define the distribution in terms of conditional probabilities:
  - E.g. \( P(Fever | Pneumonia = T) \)
  - \( P(Fever | Pneumonia = F) \)

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.**
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBC count (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: \(2*2*2*3*2 = 48\)
  - We need to define at least 47 probabilities.

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

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

  - Sum over: \(2*2*3*2 = 24\) 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 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
  \[ P(A, B \mid C) = P(A \mid C)P(B \mid C) \]
  \[ P(A \mid C, B) = P(A \mid C) \]

Alarm system example.
• Assume your house has an alarm system against burglary. You live in the seismically active area and the alarm system can get occasionally set off by an earthquake. You have two neighbors, Mary and John, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.
• We want to represent the probability distribution of events:
  – Burglary, Earthquake, Alarm, Mary calls and John calls

Causal relations

\[
\begin{align*}
\text{Burglary} & \quad \text{Earthquake} \\
\text{Alarm} & \quad \text{MaryCalls} \\
\text{JohnCalls} & \quad \text{Alarm}
\end{align*}
\]
Bayesian belief network.

1. Directed acyclic graph
   - **Nodes** = random variables
     Burglary, Earthquake, Alarm, Mary calls and John calls
   - **Links** = direct (causal) dependencies between variables.
     The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm

2. Local conditional distributions
   - relate variables and their parents
**Bayesian belief network.**

![Diagram of a Bayesian belief network with nodes for Burglary, Earthquake, Alarm, JohnCalls, and MaryCalls, and conditional probability tables for each node.]

**Full joint distribution in BBNs**

**Full joint distribution** is defined in terms of local conditional distributions (obtained via the chain rule):

\[
P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid pa(X_i))
\]

**Example:**

Assume the following assignment of values to random variables:

\[
B = T, E = T, A = T, J = T, M = F
\]

Then its probability is:

\[
P(B = T, E = T, A = T, J = T, M = F) = P(B = T)P(E = T)P(A = T \mid B = T, E = T)P(J = T \mid A = T)P(M = F \mid A = T)
\]