The ‘Asia’ (chest-clinic) example

Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, more than one of these diseases or none of them.

A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis.

The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.
### The ‘Asia’ (chest-clinic) example

Now … a patient presents with shortness-of-breath (dyspnoea) … How can the physician use available tests (X-ray) and enquiries about the patient’s history (smoking, visits to Asia) to help to diagnose which, if any, of tuberculosis, lung cancer, or bronchitis is the patient probably suffering from?

### An example from forensic genetics

DNA profiling based on STR’s (single tandem repeats) are finding many uses in forensics, for identifying suspects, deciding paternity, etc. Can we use Mendelian genetics and Bayes’ theorem to make probabilistic inference in such cases?

### Graphical model for a paternity enquiry - allowing mutation

Having observed the genotype of the child, mother and putative father, is the putative father the true father?

### Surgical rankings

- 12 hospitals carry out different numbers of a certain type of operation:
  47, 148, 119, 810, 211, 196, 148, 215, 207, 97, 256, 360 respectively.
- They are differently successful, and there are:
  0, 18, 8, 46, 8, 13, 9, 31, 14, 8, 29, 24 fatalities, respectively.

### Surgical rankings, continued

- What inference can we draw about the relative qualities of the hospitals based on these data?
- Does knowing the mortality at one hospital tell us anything at all about the other hospitals - that is, can we ‘pool’ information?

### B. Key ideas
Key ideas in exact probability calculation in complex systems

- Graphical model (usually a directed acyclic graph)
- Conditional independence graph
- Decomposability
- Probability propagation: ‘message-passing’

Let’s motivate this with some simple examples....

Directed acyclic graph (DAG)

\[ p(A,B,C) = p(A,B)p(C \mid A,B) = p(A,B)p(C \mid B) \]

\[ = p(A,B)p(B,C) \]

true for any \( A, B, C \)

Definition of \( p(C\mid B) \)

The corresponding Conditional independence graph (CIG) is

\[ p(A,B,C,D) = p(A,B)p(C \mid A,B)p(D \mid A,B,C) \]

\[ = p(A,B)p(C \mid B)p(D \mid B) \]

\[ = \frac{p(A,B)p(B,C)p(B,D)}{p(B)p(B)} \]

\[ p(A,B,C,D,E) = p(A,B)p(C,D \mid A,B)p(E \mid A,B,C,D) \]

\[ = p(A,B)p(C,D \mid B)p(E \mid C,D) \]

\[ = \frac{p(A,B)p(B,C,D)p(C,D,E)}{p(B)p(C,D)} \]

\[ p(A,B,C,D,E) = \frac{p(A,B)p(B,C,D)p(C,D,E)}{p(B)p(C,D)} \]
**Decomposability**

An important concept in processing information through undirected graphs is **decomposability**

(= graph triangulated = no chordless \( \geq 4 \)-cycles)

**Is decomposability a serious constraint?**

- How many graphs are decomposable?

<table>
<thead>
<tr>
<th>Number of vertices</th>
<th>Proportion of graphs that are decomposable</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \leq 3 )</td>
<td>all</td>
</tr>
<tr>
<td>4</td>
<td>61/64 – all but: ( \geq 80% )</td>
</tr>
<tr>
<td>6</td>
<td>( \geq 45% )</td>
</tr>
</tbody>
</table>

- Models using decomposable graphs are ‘dense’

**Is decomposability any use?**

- Maximum likelihood estimates can be computed exactly in decomposable models

\[
\hat{E}(N_{ij}) = \frac{n_{ij} n_{ji}}{n_{i+j}}
\]

- Decomposability is a key to the ‘message passing’ algorithms for probabilistic expert systems (and peeling genetic pedigrees)

**Cliques**

A **clique** is a **maximal complete subgraph**: here the cliques are \( \{1,2\}, \{2,6,7\}, \{2,3,6\}, \) and \( \{3,4,5,6\} \)
A graph is decomposable if and only if it can be represented by a junction tree (which is not unique).

The running intersection property:
For any 2 cliques C and D, C ∩ D is a subset of every node between them in the junction tree.

Non-uniqueness of junction tree

C. The works

Exact probability calculation in complex systems
0. Start with a directed acyclic graph
1. Find corresponding Conditional Independence Graph
2. Ensure decomposability
3. Probability propagation: ‘message-passing’

1. Finding the (undirected) conditional independence graph for a given DAG
   - Step 1: moralise (parents must marry)
1. Finding the (undirected) conditional independence graph for a given DAG

- Step 2: drop directions

2. Ensuring decomposability

.... triangulate

3. Probability propagation

If the distribution $p(X)$ has a decomposable CIG, then it can be written in the following potential representation form:

$$p(X) = \prod_{c \in \text{cliques}} \prod_{s \in \text{separators}} \psi(X_c | X_s)$$

the individual terms are called potentials; the representation is not unique

The potential representation can easily be initialised by

- assigning each DAG factor to (one of) the clique(s) containing $v \& \text{pa}(v)$
- setting all separator terms to 1
We can then manipulate the individual potentials, maintaining the identity
\[ p(X) = \prod_{\text{cliques } C} \psi(X_C) \prod_{\text{separators } S} \psi(X_S) \]

- first until the potentials give the clique and separator marginals,
- and subsequently so they give the marginals, conditional on given data.
- The manipulations are done by ‘message-passing’ along the branches of the junction tree.

\[ p(A, B, C) = p(A|B)p(B|C)p(C) \]

**Problem setup**

We now have a valid potential representation

\[ p(X) = \prod_{\text{cliques } C} \psi(X_C) \prod_{\text{separators } S} \psi(X_S) \]

\[ p(A, B, C) = \frac{\psi(A, B)\psi(B, C)}{\psi(B)} \]

but individual potentials are not yet marginal distributions.
We now have a valid potential representation where individual potentials are marginals:

\[
p(X) = \frac{\prod p(X_c)}{\prod_{\text{separators } S} p(X_S)}
\]

\[
p(A,B,C) = \frac{p(A,B)p(B,C)}{p(B)}
\]

We now have a valid potential representation

\[
p(X) = \prod_{\text{cliques } E} \psi(X_E)
\]

\[
p(A,B,C) = \frac{\psi(A,B)\psi(B,C)}{\psi(B)}
\]

where

\[
\psi(X_E) = p(X_E \cap \{A = 0\})
\]

for any clique or separator \(E\)
Scheduling messages

There are many valid schedules for passing messages, to ensure convergence to stability in a prescribed finite number of moves.

The easiest to describe uses an arbitrary root-clique, and first collects information from peripheral branches towards the root, and then distributes messages out again to the periphery.

D. Applications
DNA profiling based on STR's (single tandem repeats) are finding many uses in forensics, for identifying suspects, deciding paternity, etc. Can we use Mendelian genetics and Bayes’ theorem to make probabilistic inference in such cases?

**Graphical model for a paternity enquiry - neglecting mutation**

Having observed the genotype of the child, mother and putative father, is the putative father the true father?

Suppose we are looking at a gene with only 3 alleles - 10, 12 and 'x', with population frequencies 28.4%, 25.9%, 45.6% - the child is 10-12, the mother 10-10, the putative father 12-12

⇒ we’re 79.4% sure the putative father is the true father

**Graphical model for a paternity enquiry - allowing mutation**

Having observed the genotype of the child, mother and putative father, is the putative father the true father?

DNA forensics example  
(thanks to Julia Mortera)

- A blood stain is found at a crime scene
- A body is found somewhere else!
- There is a suspect
- DNA profiles on all three - crime scene sample is a ‘mixed trace’: is it a mix of the victim and the suspect?
DNA forensics in Hugin

- Disaggregate problem in terms of paternal and maternal genes of both victim and suspect.
- Assume Hardy-Weinberg equilibrium
- We have profiles on 8 STR markers - treated as independent (linkage equilibrium)

DNA forensics

The data:

<table>
<thead>
<tr>
<th>Marker</th>
<th>Victim</th>
<th>Suspect</th>
<th>Crime scene</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>VWA</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>TH01</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>TPOX</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>D5S818</td>
<td>12</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>D13S317</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>FGA</td>
<td>22</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>D7S820</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

2 of 8 markers show more than 2 alleles at crime scene ⇒ mixture of 2 or more people

DNA forensics in Hugin

DNA forensics

Population gene frequencies for D7S820 (used as ‘prior’ on ‘founder’ nodes):

<table>
<thead>
<tr>
<th>Allele</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.185</td>
</tr>
<tr>
<td>10</td>
<td>.135</td>
</tr>
<tr>
<td>11</td>
<td>.234</td>
</tr>
<tr>
<td>x</td>
<td>.233</td>
</tr>
<tr>
<td>y</td>
<td>.214</td>
</tr>
</tbody>
</table>

Results (suspect+victim vs. unknown+victim):

<table>
<thead>
<tr>
<th>Marker</th>
<th>Victim</th>
<th>Suspect</th>
<th>Crime scene</th>
<th>Likelihood ratio (sv/uv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>11.35</td>
</tr>
<tr>
<td>VWA</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>15.43</td>
</tr>
<tr>
<td>TH01</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5.44</td>
</tr>
<tr>
<td>TPOX</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>3.00</td>
</tr>
<tr>
<td>D5S818</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>14.79</td>
</tr>
<tr>
<td>D13S317</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>24.45</td>
</tr>
<tr>
<td>FGA</td>
<td>22</td>
<td>26</td>
<td>25</td>
<td>76.92</td>
</tr>
<tr>
<td>D7S820</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>4.90</td>
</tr>
</tbody>
</table>

overall 3.93 × 10^8
Surgical rankings

- 12 hospitals carry out different numbers of a certain type of operation: 47, 148, 119, 810, 211, 196, 215, 207, 97, 256, 360 respectively.
- They are differently successful, and there are: 0, 18, 8, 46, 8, 13, 9, 31, 14, 8, 29, 24 fatalities, respectively.

Surgical rankings, continued

- What inference can we draw about the relative qualities of the hospitals based on these data?
- A natural model is to say the number of deaths $y_i$ in hospital $i$ has a Binomial distribution $y_i \sim Bin(n_i, p_i)$ where the $n_i$ are the numbers of operations, and it is the $p_i$ that we want to make inference about.

Surgical rankings, continued

- How to model the $p_i$?
- We do not want to assume they are all the same.
- But they are not necessarily `completely different'.
- In a Bayesian approach, we can say that the $p_i$ are random variables, drawn from a common distribution.

Specifically, we could take

$$\log \frac{p_i}{1-p_i} \sim N(\theta, \sigma^2)$$

If $\theta$ and $\sigma^2$ are fixed numbers, then inference about $p_i$ only depends on $y_i$, (and $n_i$, $\theta$ and $\sigma^2$).

Surgical rankings, continued

- But don't you think that knowing that $p_1 = 0.08$, say, would tell you something about $p_2$?
- Putting prior distributions on $\theta$ and $\sigma^2$ allows ‘borrowing strength' between data from different hospitals.
Surgical rankings - simplified
3 hospitals, $p$ discrete, only one hyperparameter

The 'Asia' (chest-clinic) example
Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, more than one of these diseases or none of them. A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.
The ‘Asia’ (chest-clinic) example

Now … a patient presents with shortness-of-breath (dyspnoea) …. How can the physician use available tests (X-ray) and enquiries about the patient’s history (smoking, visits to Asia) to help to diagnose which, if any, of tuberculosis, lung cancer, or bronchitis is the patient probably suffering from?

E. Proofs

Factorisation of joint distribution, forming potential representation, when graph is decomposable

Decomposability

The following are equivalent

• $G$ is decomposable
• $G$ is triangulated (or chordal)
• The cliques of $G$ may be ‘perfectly numbered’ to satisfy the running intersection property

\[ C_i \cap \bigcup_{j<i} C_j \subseteq C_i \forall i = 2,3,\ldots,k \]

where \( i' \in \{1,2,\ldots,i-1\} \)

Decomposability

$G$ is decomposable means that either

• $G$ is complete, or
• $G$ admits a proper decomposition $(A,B,C)$, that is:
  – $B$ separates $A$ and $C$
  – $B$ is complete, $A$ and $C$ are non-empty
  – the subgraphs $G_{A,B}$ and $G_{B,C}$ are decomposable
Decomposability

$G$ is triangulated or chordal means that

- $G$ has no loops of 4 or more vertices without a chord

The running intersection property

$$C_i \cap \bigcup_{j<i} C_j \subseteq C_i \forall i = 2,3,\ldots,k$$

is what allows the construction of the junction tree and the possibility of probability propagation.

The junction tree

For $i=2,3,\ldots,k$, join $C_i$ to $C_{i-1}$, labelling the edge by $S_i$, and $S_i \subseteq C_{i-1} \forall i = 2,3,\ldots,k$

A decomposable graph and (one of) its junction tree(s)

Decomposability

In

$$C_i \cap \bigcup_{j<i} C_j \subseteq C_i \forall i = 2,3,\ldots,k$$

let

$$S_i = C_i \cap \bigcup_{j<i} C_j$$
$$R_i = C_i \setminus S_i$$
$$H_{i-1} = \bigcup_{j<i} C_j$$

then

$$S_i = C_i \cap H_{i-1} \subseteq C_i \forall i = 2,3,\ldots,k$$

$S_i$ separates $R_i$ & $H_{i-1}$
Factorisation of joint distribution

\[ H_{i-1} = \bigcup_{j<i} C_j, \]

Recall \( H_{i-1} = \bigcup_{j<i} C_j \), then

\[ p(V) = p(H_i) p(C_2 \setminus H_1 \mid H_i) \times \]

\[ p(C_3 \setminus H_2 \mid H_2) \ldots p(C_k \setminus H_{k-1} \mid H_{k-1}) \]

but the typical factor is

\[ p(C_i \setminus H_{i-1} \mid H_{i-1}) = p(R_i \mid H_{i-1}) \]

\[ = p(R_i \mid S_i) = \frac{p(R_i, S_i)}{p(S_i)} = \frac{p(C_i)}{p(S_i)} \]

\[ p(C_i) \]

\[ \prod_{i=2}^{k} p(S_i) \]

as required

E. Proofs

The collect/distribute schedule ensures equilibrium in message-passing

Scheduling messages

There are many valid schedules for passing messages, to ensure convergence to stability in a prescribed finite number of moves.

The easiest to describe uses an arbitrary root-clique, and first collects information from peripheral branches towards the root, and then distributes messages out again to the periphery.
Consider a single edge of the junction tree

(I, J and K may be vectors)

- Edge is in equilibrium if J table is equal to J marginal in both IJ and JK tables
- Tree is in equilibrium if every edge is


Messages passed from JK to root and back to JK

As a result, JK table gets multiplied by a term indexed by \((j,k)\) – but not \(i\)

Messages passed from IJ back to leaves

IJ, J and JK tables are not changed again
Final tables

$$\left( \sum_{i} c_{i} d_{i} \right) \frac{a_{i}}{b_{i}} \left( \sum_{i} c_{i} d_{i} \right) \frac{a_{i}}{b_{i}} \left( c_{i} d_{i} \right) \frac{a_{i}}{b_{i}}$$

- satisfy equilibrium conditions

Software

- The HUGIN system: freeware version (Hugin Lite 5.7):
  http://www.stats.bris.ac.uk/~peter/Hugin57.zip
- Grappa (suite of R functions)
  http://www.stats.bris.ac.uk/~peter/Grappa

Module outline

- Information, uncertainty and probability
- Motivating examples
- Graphical models
- Probability propagation
- The HUGIN system