Decision Support Models 2010/2011
Bayesian Nets

Lecture topics:
Bayesian belief networks; Holmes & Watson example
Software for implementing belief networks
Case Study: Chest Clinic

References:
Jensen, F.V. (1996), An Introduction to Bayesian Networks, UCL Press, London. (Ch. 2)
Breese J. (Microsoft Corp.), Koller, D. (Stanford University):
http://robotics.stanford.edu/~koller/BNtut/
Norsys Software Corp. (NETICA soft.): http://www.norsys.com
A Taxonomy of Decision Models
(In Decision Analysis in the 1990s - L.D. Phillips)

Problem dominated by

Uncertainty

Multiple Objectives

EXTEND conversation
- Event tree
- Fault tree
- Influence diagram

REVISE opinion
- Bayesian nets

CHOOSE option
- Payoff matrix
- Decision tree

EVALUATE options
- Multi-criteria decision analysis

ALLOCATE resources
- Multi-criteria commons dilemma

NEGOTIATE
- Multi-criteria bargaining analysis

SEPARATE into components
- Credence decomposition
- Risk analysis
Remember: Rules of Probability

Product rule:
\[ P(\text{Favourable, Oil}) = P(\text{Favourable|Oil}) \cdot P(\text{Oil}) = P(\text{Oil|Favourable}) \cdot P(\text{Favourable}) \]

Marginalisation:
\[ P(\text{Favourable}) = P(\text{Favourable|Oil}) \cdot P(\text{Oil}) + P(\text{Favourable|Dry}) \cdot P(\text{Dry}) \]

Bayes’ Rule (1763):
\[ P(\text{Oil|Favourable}) = \frac{P(\text{Favourable|Oil}) \cdot P(\text{Oil})}{P(\text{Favourable})} \]
\[ P(\text{Oil|Favourable}) = \frac{P(\text{Favourable|Oil}) \cdot P(\text{Oil})}{P(\text{Favourable|Oil}) \cdot P(\text{Oil}) + P(\text{Favourable|Dry}) \cdot P(\text{Dry})} \]
where \( P(H \mid E) \) denotes the probability of hypothesis \( H \) conditioned on the evidence \( E \).

Bayes’ rule provides an explicit relation for the degree of believe we accord a hypothesis \( H \), in light of evidence \( E \).

Bayes’ Rule is useful in contexts where probabilities are more easily obtained in one inferential direction than another.
By providing the flexibility to reason probabilistically in either the *causal* or the *diagnostic* directions, Bayes’ Rule allows agents to *assert* beliefs in forms that are *compatible* with the way they actually reason about the process(es) or phenomena of interest.

**A bayesian network**

A particular type of Influence Diagram. It *contains only* chance (and deterministic) *nodes*.

**Collins Case (Edwards, 1991)**

[Diagram of a Bayesian network with nodes like "Brooks is mugged by Caucasian female" and "Mugger has blond pony tail" connected by arrows representing causal relationships.]
• Lots of variables
• Lots of complicated relationships
• Need to build an inference structure

A classic example of the use of belief networks is in the medical domain. Here each new patient typically corresponds to a new "case", and the problem is to diagnose the patient (i.e. find beliefs for the unmeasurable disease variables), predict what is going to happen to the patient, or find an optimal prescription, given the values of observable variables (symptoms). A doctor may be the expert used to define the structure of the network, and provide the initial relations between variables (often in the form of conditional probabilities), based on his medical training and experience with previous cases. Then the network probabilities may be fine-tuned by using statistics from previous cases, and from new cases as they arrive.
Bayesian (or Belief, or Probabilistic Causal) Networks

A BN is composed of a set of nodes representing variables of interest, connected by links to indicate dependencies, and containing information about the relationships between the nodes (often in the form of conditional probabilities). Its uses include prediction and diagnosis.

A BN provides a complete probabilistic description of a particular system, i.e., completely specifies a joint probability distribution on the kinds of distinctions represented by the network.

Remember: Conditional \( P(X = x \mid Y = y) \)

Probability that \( X=x \) given we know that \( Y=y \).

Joint \( P(x, y) \equiv P(X = x \land Y = y) \)

Probability that both \( X = x \) and \( Y = y \).
A BN consists of the following:

- A set of variables and a set of directed edges between variables.
- Each variable has a finite set of states.
- The variables together with the directed edges form a directed acyclic graph.
- To each variable $A$ with parents $B_1, \ldots, B_n$ there is attached a conditional probability table $P(A \mid B_1, \ldots, B_n)$. 

### (FORMAL) DEFINITION OF A BN (Jansen, 1996)

A BN consists of the following:

- A set of variables and a set of directed edges between variables.
- Each variable has a finite set of states.
- The variables together with the directed edges form a directed acyclic graph.
- To each variable $A$ with parents $B_1, \ldots, B_n$ there is attached a conditional probability table $P(A \mid B_1, \ldots, B_n)$.
Holmes & Watson example

Police Inspector Smith is impatiently waiting the arrival of Mr Holmes and Dr Watson; they are late and Inspector Smith has another important appointment (lunch). Looking out of the window he wonders whether the roads are icy. Both are notoriously bad drivers, so if the roads are icy they may crash.

His secretary enters and tells him that Dr Watson has had a car accident, “Watson? OK. It could be worse… icy roads! Then Holmes has most probably crashed too. I’ll go for lunch now.”

“Icy roads?” the secretary replies, “It is far from being that cold, and furthermore all the roads are salted.” Inspector Smith is relieved. “Bad luck for Watson. Let us give Holmes ten minutes more.”
Police Inspector Smith is impatiently waiting the arrival of Mr Holmes and Dr Watson; they are late and Inspector Smith has another important appointment (lunch). Looking out of the window he wonders whether the roads are icy. Both are notoriously bad drivers, so if the roads are icy they may crash.

**Bayesian net**

<table>
<thead>
<tr>
<th>Ice</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>80.000</td>
<td>20.000</td>
</tr>
<tr>
<td>False</td>
<td>50.000</td>
<td>50.000</td>
</tr>
</tbody>
</table>

Both are notoriously bad drivers

<table>
<thead>
<tr>
<th>Ice</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Roads are icy they may crash

<table>
<thead>
<tr>
<th>Ice</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>80.0</td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>20.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ice</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>30.0</td>
<td></td>
</tr>
</tbody>
</table>
if the roads are icy they may crash
if the roads are icy they may crash

His secretary enters and tells him that Dr Watson has had a car accident, “Watson? OK. It could be worse… icy roads! Then Holmes has most probably crashed too. I’ll go for lunch now.”
Then Holmes has most probably crashed too

“Icy roads?”, the secretary replies, “It is far from being that cold, and furthermore all the roads are salted.” Inspector Smith is relieved. “Bad luck for Watson. Let us give Holmes ten minutes more.”

<table>
<thead>
<tr>
<th>Ice</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>80.000</td>
<td>20.000</td>
</tr>
<tr>
<td>False</td>
<td>50.000</td>
<td>50.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Holmes</th>
<th>Accident</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>50.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ice</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>100</td>
</tr>
</tbody>
</table>
Qualitative properties of Bayes nets

- Inference can follow arrow direction
  - Knowing T tells you something about H via I
- Inference can run against arrow direction
  - Knowing H tells you something about T via I
- Both patterns of inference are broken if I know I
• Inference can also work up along an arrow and then back down
  – Knowing H can tell me something about W, supposing I don’t know I

• This is known as *conditional independence*

  • H and W (or H and T) are independent if I know the state of I (otherwise they are dependent)
• What about working down and then up?

• If I don’t know the state of L, R and S are independent

• But if I know L is the case, R and S become dependent
  – Knowing S can explain away L and thus make R less likely

• This pattern of inference is enabled rather than broken by knowledge of intermediate variable
d-separation

Definition. Two variables A and B in a Bayes net are *d-separated* if for all paths between A and B there is an intermediate variable V such that

- the connection is serial or diverging and the state of V is known
- the connection is converging and neither V nor any of V’s descendants have received evidence
Significance of d-separation

• It’s been suggested that d-separation represents a basic property which any automation of reasoning under uncertainty must obey.

• Equally, it could be argued that this is such a non-intuitive property that it’s further evidence that evolution hasn’t equipped us with great natural tools for probabilistic inference.
This belief network is also known as "Asia". It is a toy medical diagnosis example from: Lauritzen, Steffen L. and David J. Spiegelhalter (1988), “Local computations with probabilities on graphical structures and their application to expert systems”, *J. Royal Statistics Society B*, 50(2), 157-194.
It is a simplified version of a network that could be used to diagnose patients arriving at a clinic. Each node in the network corresponds to some condition of the patient, for example, "Visit to Asia" indicates whether the patient recently visited Asia. To diagnose a patient, values are entered for nodes when they are known. Netica then automatically re-calculates the probabilities for all the other nodes, based on the relationships between them. The links between the nodes indicate how the relationships between the nodes are structured.
The two top nodes are for predispositions which influence the likelihood of the diseases. Those diseases appear in the row below them. At the bottom are symptoms of the diseases. To a large degree, the links of the network correspond to causation. This is a common structure for diagnostic networks: predisposition nodes at the top, with links to nodes representing internal conditions and failure states, which in turn have links to nodes for observables. Often there are many layers of nodes representing internal conditions, with links between them representing their complex inter-relationships.
Probabilistic relation of "Lung Cancer" with Smoking

Functional dependence of "Tuberculosis or Cancer" on Tuberculosis and Lung Cancer.
Suppose we want to "diagnose" a new patient. When she first enters the clinic, without having any information about her, we believe she has lung cancer with a probability of 5.5% (the number may be higher than that for the general population, because something has led her to the chest clinic).
Finding - She has an abnormal x-ray

All the probability numbers and bars changed to take into account the finding. Now the probability that she has lung cancer has increased from 5.5% to 48.9%.
The probability of lung cancer decreases from 48.9% to 37.1%, because the abnormal XRay is partially explained away by a greater chance of Tuberculosis (which she could catch in Asia).

New Finding: She has made a visit to Asia recently.
A new patient has just walked in:
remove all the findings
THE COLLINS CASE  (Edwards, 1991)

Janet and Malcolm Collins were convicted by a Los Angeles jury of second-degree robbery. Malcolm appealed his case to the Supreme Court of California on several grounds. In the opinion of the court, the crucial one was “that the introduction of evidence pertaining to the mathematical theory of probability and the use of the same by the prosecution during the trial was error prejudicial to the defendant.” The California Supreme Court agreed and reversed the conviction. The court described the facts of the case as follows:

On June 18, 1964, about 11:30 A.M. Mrs. Juanita Brooks, who had been shopping, was walking home along an alley in the San Pedro area of the City of Los Angeles… As she stooped down to pick up an empty carton, she was suddenly pushed to the ground by a person whom she neither saw nor heard approach… She managed to look up and saw a young woman running from the scene. According to Mrs. Brooks the latter appeared to weigh about 145 pounds, was wearing “something dark,” and had hair “between a dark blond and a light blond,”… [H]er purse, containing between $35 and $40, was missing.
THE COLLINS CASE (cont.)

About the same time…, John Bass, who lived on the street at the end of the alley, …[heard] “a lot of crying and screaming” coming from the alley. As he looked in that direction, he saw a woman run out of the alley and enter a yellow automobile parked across the street from him… The car… passed within six feet of Bass… It was being driven by a male Negro, wearing a moustache and beard… Bass described the woman who ran from the alley as a Caucasian, slightly over five feet tall, of ordinary build, with her hair in a dark blond ponytail, and wearing dark clothing.
### Table 1: Probabilities Used by the Prosecutor in the Collins Case

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partly yellow automobile</td>
<td>0.1</td>
</tr>
<tr>
<td>Man with moustache</td>
<td>0.25</td>
</tr>
<tr>
<td>Girl with ponytail</td>
<td>0.1</td>
</tr>
<tr>
<td>Girl with blond hair</td>
<td>0.333</td>
</tr>
<tr>
<td>Black man with beard</td>
<td>0.1</td>
</tr>
<tr>
<td>Interracial couple in car</td>
<td>0.001</td>
</tr>
<tr>
<td>Table 2: PROBABILITIES IN THE COLLINS CASE INFLUENCE DIAGRAM</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Node 1: BROOKS IS MUGGED BY CAUCASIAN FEMALE</strong></td>
<td></td>
</tr>
<tr>
<td>OUTCOMES</td>
<td>PROBABILITIES</td>
</tr>
<tr>
<td>Yes, Janet Collins</td>
<td>.0001</td>
</tr>
<tr>
<td>Other Caucasian Female</td>
<td>.9999</td>
</tr>
<tr>
<td><strong>Node 2: MUGGER HAS BLOND PONYTAIL</strong></td>
<td></td>
</tr>
<tr>
<td>CONDITION OF NODE 1</td>
<td>OUTCOMES</td>
</tr>
<tr>
<td>Yes, Janet Collins</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>Other Caucasian Female</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>Node 3: GETAWAY CAR IS YELLOW</strong></td>
<td></td>
</tr>
<tr>
<td>CONDITION OF NODE 1</td>
<td>OUTCOMES</td>
</tr>
<tr>
<td>Yes, Janet Collins</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>Other Caucasian Female</td>
<td>0.100</td>
</tr>
<tr>
<td><strong>Node 4: DRIVER OF CAR IS BLACK</strong></td>
<td></td>
</tr>
<tr>
<td>CONDITION OF NODE 1</td>
<td>OUTCOMES</td>
</tr>
<tr>
<td>Yes, Janet Collins</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>Other Caucasian Female</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Node 5: DRIVER FACIAL FOLIAGE</strong></td>
<td></td>
</tr>
<tr>
<td>IF NODE 1 IS &quot;Yes, Janet Collins&quot; AND NODE 4 IS...</td>
<td>OUTCOMES</td>
</tr>
<tr>
<td>Yes (driver Black)</td>
<td>Beard and moustache</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>No (driver not Black)</td>
<td>0.025</td>
</tr>
<tr>
<td>IF NODE 1 IS &quot;Other Caucasian Female&quot; AND NODE 4 IS...</td>
<td></td>
</tr>
<tr>
<td>Yes (driver Black)</td>
<td>0.025</td>
</tr>
<tr>
<td>No (driver not Black)</td>
<td>0.025</td>
</tr>
</tbody>
</table>
### Table 3: Probabilities, Likelihood Ratios, and Weights of Evidence in the Collins Case

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Prior Probability</th>
<th>Posterior Probability</th>
<th>Likelihood Ratio (LR)</th>
<th>Weight of Evidence (log LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blond with ponytail</td>
<td>0.0001</td>
<td>0.0030</td>
<td>30,000</td>
<td>+1.4771</td>
</tr>
<tr>
<td>Yellow car</td>
<td>0.0001</td>
<td>0.0010</td>
<td>9.990</td>
<td>+0.9996</td>
</tr>
<tr>
<td>Black with moustache and beard</td>
<td>0.0001</td>
<td>0.7997</td>
<td>39,920.000</td>
<td>+4.6012</td>
</tr>
<tr>
<td>All three aggregated</td>
<td>0.0001</td>
<td>0.9992</td>
<td>11,964,036.000</td>
<td>+7.0779</td>
</tr>
<tr>
<td>Blackness alone</td>
<td>0.0001</td>
<td>0.0908</td>
<td>999,000</td>
<td>+2.9996</td>
</tr>
<tr>
<td>Facial foliage alone</td>
<td>0.0001</td>
<td>0.0040</td>
<td>39.921</td>
<td>+1.6012</td>
</tr>
<tr>
<td>Blackness, given facial foliage</td>
<td>0.0040</td>
<td>0.7997</td>
<td>999.975</td>
<td>+3.0000</td>
</tr>
<tr>
<td>Facial foliage, given blackness</td>
<td>0.0908</td>
<td>0.7997</td>
<td>39.960</td>
<td>+1.6016</td>
</tr>
</tbody>
</table>
DISCUSSION QUESTIONS

• Role of probability in the courtroom/legal litigation?
  — Normative, descriptive, and prescriptive interactions?
• Suppose that in the Collins Case you are hired as an expert consultant for the defence. How might you attack or counter the analysis that we discussed today?
Structuring

Network structure corresponding to “causality” is usually good.

Extending the conversation.

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Independence

Age and Gender are independent.

\[ P(A, G) = P(G) P(A) \]
\[ P(A|G) = P(A) \quad A \perp G \]
\[ P(G|A) = P(G) \quad G \perp A \]

\[ P(A, G) = P(G|A) P(A) = P(G) P(A) \]
\[ P(A, G) = P(A|G) P(G) = P(A) P(G) \]
Evidence on $A$ (or $G$) will influence the certainty of $S$. Similarly, evidence on $C$ will influence the certainty on $A$ (and $G$) through $S$. However, if the state of $S$ is known, then $A$ ($G$) and $C$ become independent.

We say that $A$ ($G$) and $C$ are \textit{d-separated} given $B$.

\[ P(C|A,G,S) = P(C|S) \quad C \perp A,G \mid S \]

\textbf{Serial connection.} If $S$ is instantiated (i.e., its state is known) it blocks communication between its parents and children.
More Conditional Independence: Naïve Bayes

When nothing is known about the state of Cancer:

*Serum Calcium and Lung Tumor are dependent*

However, when we know the state of Cancer:

*Serum Calcium is independent of Lung Tumor, given Cancer*

\[ P(L|SC,C) = P(L|C) \]

**Diverging connection.** If C is instantiated, it blocks communication between its children.
More Conditional Independence: Explaining Away

When nothing is known about the state of Cancer:

*Exposure to Toxics* and *Smoking* are independent

\[ E \perp S \]

However, when we have information about Cancer:

*Exposure to Toxics* is dependent on *Smoking*, given Cancer

\[
P(E = \text{heavy} \mid C = \text{malignant}) > \]

\[
P(E = \text{heavy} \mid C = \text{malignant}, S=\text{heavy})
\]

**Converging connection.**
If \( C \) is instantiated, it opens communication between its parents
Put it all together

\[
P(A, G, E, S, C, L, SC) = P(A) \cdot P(G) \cdot \\
P(E \mid A) \cdot P(S \mid A, G) \cdot \\
P(C \mid E, S) \cdot \\
P(SC \mid C) \cdot P(L \mid C)
\]
The joint probability distribution is the product of all conditional probabilities:

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

where: \( Pa_i = \text{parents}(X_i) \)
Conditional Independence

A variable (node) is conditionally independent of its non-descendants given its parents.

\[
\begin{align*}
\text{Age} & \quad \text{Gender} \\
\text{Exposure to Toxics} & \quad \text{Smoking} \\
\text{Cancer} & \quad \text{Serum Calcium} \\
& \quad \text{Lung Tumor}
\end{align*}
\]

\text{Cancer} \text{ is independent of Age and Gender given Exposure to Toxics and Smoking.}

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.
Two variables are d-separated if for all paths between them there is an intermediate variable $V$ such that either:

- the connection is serial or diverging and the state of $V$ is known or
- the connection is converging, and neither $V$ nor any of $V$'s descendents have received evidence.

Independence and Graph Separation

- Given a set of observations, is one set of variables dependent on another set?
- Observing effects can induce dependencies.
- $d$-separation (Pearl 1988) allows us to check conditional independence graphically.
The nodes are then connected up with directed links. If there is a link from node A to node B, then node A is sometimes called the parent, and node B the child (of course, B could be the parent of another node). Usually a link from node A to node B indicates that A causes B, that A partially causes or predisposes B, that B is an imperfect observation of A, that A and B are functionally related, or that A and B are statistically correlated. The precise definition of a link is based on conditional independence, and is explained in detail in a reference like Neapolitan90 or Pearl88. However, most people seem to intuitively grasp the notion of links, and use them effectively without concentrating on the precise definition.

Finally, probabilistic relations are provided for each node, which express the probabilities of that node taking on each of its values, conditioned on the values of its parent nodes. Some nodes may have a deterministic relation, which means that the value of the node is given as a direct function of the parent node values.

After the belief network is constructed, it may be applied to a particular case. For each variable you know the value of, you enter that value into its node as a finding (also known as "evidence"). Then Netica does probabilistic inference to find beliefs for all the other variables. Suppose one of the nodes corresponds to the variable "Temperature", and it can take on the values cold, medium and hot. Then an example belief for temperature could be: [cold - 0.1, medium - 0.6, hot - 0.3], indicating the subjective probabilities that the temperature is cold, medium or hot.
Probabilistic inference is the process of calculating new beliefs for a set of variables, given some findings. Technically speaking, it is the process of finding a posterior distribution, given a prior distribution, a model and some observations. BNs do probabilistic inference by belief updating.

• The belief of a node is the set of probabilities (one for each of its possible states), taking into account the currently entered findings by using the knowledge encoded in the BN. Technically speaking, it is the marginal posterior probability distribution of the node, given the findings and the belief network model.

• A finding (or “evidence”) is a value for one of the nodes (i.e. variables) of a BN when it is applied to a particular situation.

Probabilistic inference only results in a set of beliefs at each node; it does not change the network (knowledge base) at all. If the findings that have been entered are a true example that might give some indication of cases which will be seen in the future, you may think that they should change the knowledge base a little bit as well, so that next time it is used, its conditional probabilities more accurately reflect the real world. To achieve this you would also do probability revision.
Predictive Inference

How likely are elderly males to get malignant cancers?

\[ P(C = \text{malignant} \mid \text{Age} > 60, \text{Gender} = \text{male}) \]
Combined

How likely is an elderly male patient with high Serum Calcium to have malignant cancer?

\[ P(C=\text{malignant} \mid \text{Age}>60, \quad \text{Gender= male, Serum Calcium = high}) \]
Explaining away

- If we see a lung tumor, the probability of heavy smoking and of exposure to toxiics both go up.
- If we then observe heavy smoking, the probability of exposure to toxiics goes back down.
Belief updating is the process of finding new beliefs for the nodes of a BN to account for the findings that are currently known. It is a form of probabilistic inference. During belief updating the BN model (in particular, the conditional probability relations between the nodes) is not modified at all; for that probability revision is used.

Probability revision is the process of adjusting the conditional probability relations of a belief network to account for a new case (i.e. set of findings), or more often, for a new set of cases. It is a form of parameter learning.

Parameter learning is the automatic learning of the specific relationships nodes have with their parents using case data, once it has already been determined which nodes are the parents of each node. These relationships are usually in the form of conditional probabilities, or the parameters of a conditional probability equation.