Faculty of Life Sciences

Introduction to Bayesian Networks

Advanced Herd Management
28th of september 2009

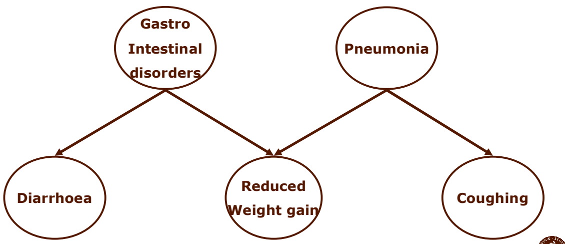
Tina Birk Jensen

What is it all about?

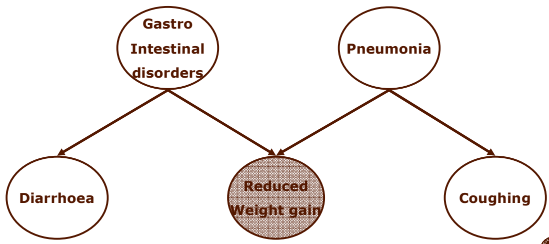
Method to reasoning under uncertainty

→ Where we reason using probabilities

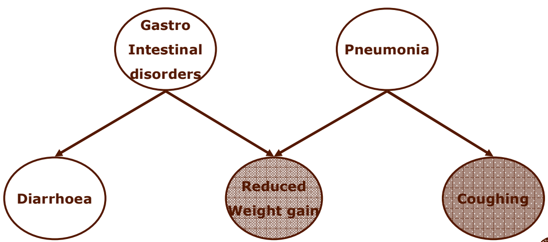
Reasoning under uncertainty: An example



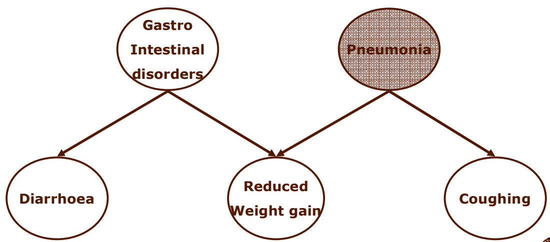
Reasoning under uncertainty: An example



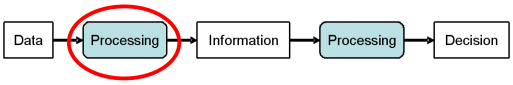
Reasoning under uncertainty: An example



Reasoning under uncertainty: An example



Where is Bayesian networks placed in AHM?



Text books/literature

1. Bayesian Networks and Decision Graphs

A *general* textbook on Bayesian networks and decision graphs.

Written by professor Finn Verner Jensen from Ålborg University – one of the leading research centers for Bayesian networks.

2. Bayesian Networks without Tears

Article written by Eugene Charniak

Software

Esthauge LIMID Software System

www.esthauge.dk

You can download the software from the course homepage

Outline

Today (28th of september)
General introduction to Bayesian networks:

- What is a Bayesian network?
- Exercise 11.1
- Transmission of evidence
- Exercise 11.2

Tuesday (29th of september)
Building Bayesian network models

Friday (2nd of October)
Case example

Bayesian networks (in general)

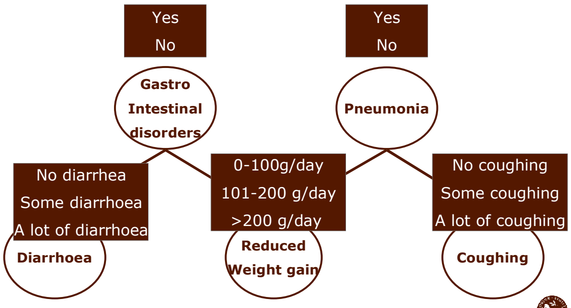
- Graphical model with some restrictions (next slide)
- Basically a static method ("here and now" imagine)
- A static version of data filtering
- All parameters are probabilities

Bayesian networks - definition

Bayesian networks consist of:

- A set of variables and a set of directed edges between variables
- Each variable has a finite set of mutually exclusive states

”Each variable has a definte set of mutually excusive states”

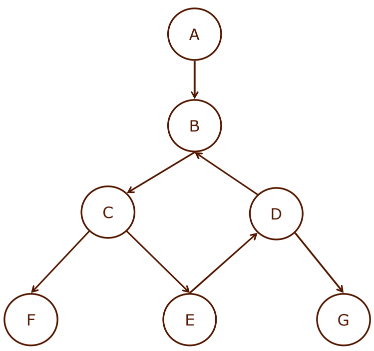


Bayesian networks - definition

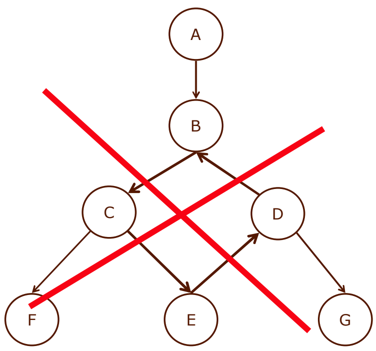
Bayesian networks consist of:

- A set of variables and a set of directed edges between variables
- Each variable has a finite set of mutually exclusive states
- The variables and directed edges form a directed acyclic graph

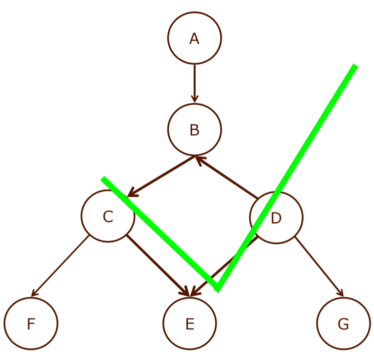
”The variables and directed edges form a directed acyclic graph”



”The variables and directed edges form a directed acyclic graph”



”The variables and directed edges form a directed acyclic graph”

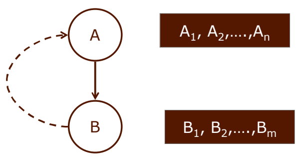


Bayesian networks - definition

Bayesian networks consist of:

- A set of variables and a set of directed edges between variables
- Each variable has a finite set of mutually exclusive states
- The variables and directed edges form a directed acyclic graph
- To each variables A with parents B_1, B_2 to B_n there is attached the probability table $P(A| B_1, B_2 .... B_n)$

Baye's Theorem

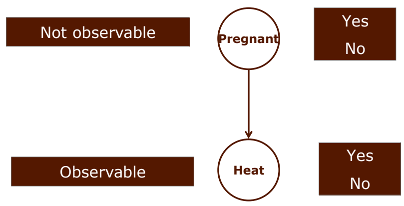


$$P(A_i \mid B_j) = \frac{P(B_j \mid A_i)P(A_i)}{P(B_j \mid A_1)p(A_1) + P(B_j \mid A_2)P(A_2) + \ldots + P(B_j \mid A_n)P(A_n)}$$

$$P(B_j) = \sum_{k=1}^{n} P(B_j \mid A_k)P(A_k)$$

Now an example!!

A small Bayesian network: Pregnancy and heat detection in cows



Pregnancy and heat detection in cows

What is the probability that a farmer ***observes*** a particular cow in heat during a 3-week period?

- P(*Heat* = "yes") = *a*
- P(*Heat* = "no") = *b*
- *a* + *b* = 1 (no other options)

What is the probability that the cow is pregnant?

- P(*Pregnant* = "yes") = *c*
- P(*Pregnant* = "no") = *d*
- *c* + *d* = 1 (no other options)

Conditional probabilities

Now, assume that the cow is pregnant.
What is the conditional probability that the farmer observes it in heat?

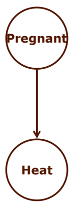
- P(*Heat* = "yes" | *Pregnant* = "yes") = a_{p+}
- P(*Heat* = "no" | *Pregnant* = "yes") = b_{p+}
- Again, $a_{p+} + b_{p+} = 1$

Now, assume that the cow is *not* pregnant.
Accordingly:

- P(*Heat* = "yes" | *Pregnant* = "no") = a_{p-}
- P(*Heat* = "no" | *Pregnant* = "no") = b_{p-}
- Again, $a_{p-} + b_{p-} = 1$

Each value of *Pregnant* defines a full probability distribution for *Heat*. Such a distribution is called ***conditional***

A small Bayesian net



***Pregnant* = "yes"**	***Pregnant* = "no"**
c = 0.5	*d* = 0.5

	***Heat* = "yes"**	***Heat* = "no"**
***Pregnant* ="yes"**	a_{p+} = 0.02	b_{p+} = 0.98
***Pregnant* = "no"**	a_{p-} = 0.60	b_{p-} = 0.40

Let us build the net!

Experience with the net: Evidence

By entering information on an observed value of *Heat* we can revise our belief in the value of the ***unobservable*** variable *Pregnant.*

The observed value of a variable is called ***evidence.***

The revision of beliefs is done by use of ***Baye's Theorem***:

$$P(A_i \mid B_j) = \frac{P(B_j \mid A_i)P(A_i)}{P(B_j \mid A_1)p(A_1) + P(B_j \mid A_2)P(A_2) + \ldots + P(B_j \mid A_n)P(A_n)}$$

Baye's Theorem for our net

How do we use Bayes formula to calculate:
P(Pregnant="yes"|Heat="yes")

$$P(A_i \mid B_j) = \frac{P(B_j \mid A_i)P(A_i)}{P(B_j \mid A_1)p(A_1) + P(B_j \mid A_2)P(A_2) + \ldots.. + P(B_j \mid A_n)P(A_n)}$$

$$\mathrm{P}(Pregnant = \text{yes} \mid Heat = \text{yes}) =$$

$$\frac{\mathrm{P}(Heat{=}\text{yes} \mid Pregnant{=}\text{yes})\mathrm{P}(Pregnant{=}\text{yes})}{\mathrm{P}(Heat{=}\text{yes} \mid Pregnant{=}\text{yes})\mathrm{P}(Pregnant{=}\text{yes}) + \mathrm{P}(Heat{=}\text{yes} \mid Pregnant{=}\text{no})\mathrm{P}(Pregnant{=}\text{no})} =$$

$$\frac{a_{p+}c}{a_{p+}c + a_{p-}d} = \frac{0.02 \times 0.5}{0.02 \times 0.5 + 0.6 \times 0.5} = 0.032258$$

Pregnant = "yes"	Pregnant = "no"
c = 0.5	d = 0.5

	Heat = "yes"	Heat = "no"
Pregnant ="yes"	a_{p+} = 0.02	b_{p+} = 0.98
Pregnant = "no"	a_{p-} = 0.60	b_{p-} = 0.40

Baye's Theorem for our net

How do we use Bayes formula to calculate:
P(Pregnant="yes"|Heat="no")

$$\mathrm{P}(Pregnant = \text{yes} \mid Heat = \text{no}) =$$

$$\frac{\mathrm{P}(Heat{=}\text{no} \mid Pregnant{=}\text{yes})\mathrm{P}(Pregnant{=}\text{yes})}{\mathrm{P}(Heat{=}\text{no} \mid Pregnant{=}\text{yes})\mathrm{P}(Pregnant{=}\text{yes}) + \mathrm{P}(Heat{=}\text{no} \mid Pregnant{=}\text{no})\mathrm{P}(Pregnant{=}\text{no})} =$$

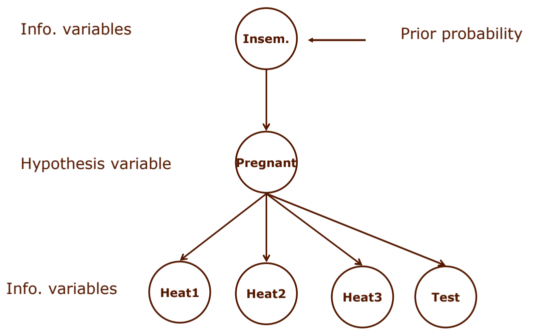
$$\frac{b_{p+}c}{b_{p+}c + b_{p-}d} = \frac{0.98 \times 0.5}{0.98 \times 0.5 + 0.4 \times 0.5} = 0.7101449$$

Pregnant = "yes"	Pregnant = "no"
c = 0.5	d = 0.5

	Heat = "yes"	Heat = "no"
Pregnant ="yes"	a_{p+} = 0.02	b_{p+} = 0.98
Pregnant = "no"	a_{p-} = 0.60	b_{p-} = 0.40

Now time for exercise 11.1!

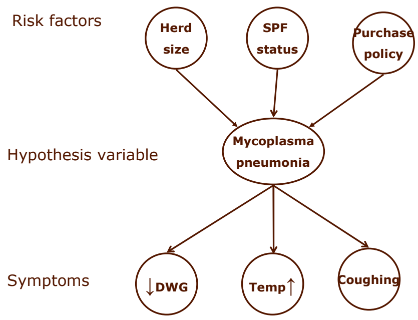
Extension of the net



Advantages of Bayesian networks

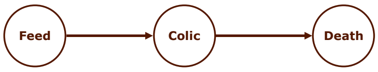
- Consistent combination of information from various sources
- Can estimate certainties for the values of variables that are not observable (or very costly to observe). These variables are called "hypothesis variables".
- These estimates are obtained by entering evidence in "information variables" that
  - Influence the hypothesis variable
  - Depend on the hypothesis variable

Herd diagnostics



Transmission of evidence

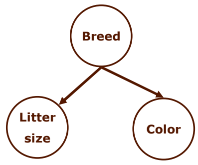
Serial connections



- If ”Colic” is observed, there will be no connection between ”Feed” and ”Death”
- ”Feed” and ”Death” are d-separated given ”Colic”
- Evidence may be transmitted through a serial connection unless, the state of the intermediate variable is known

Transmission of evidence

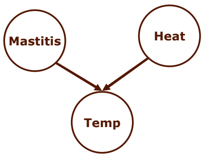
Diverging connection



- If ”Breed” is observed, there will be no influence of ”Color” on ”Litter size”
- ”Litter size” and ”Color” are d-separated given ”Breed”
- Evidence may be transmitted through a diverging connection unless, the state of the intermediate variable is known

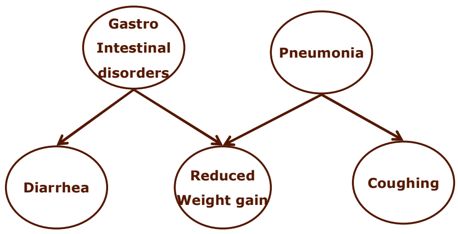
Transmission of evidence

Converging connection

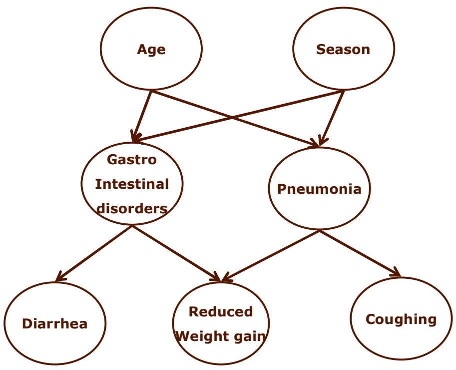


- If ”Temp” is observed, the information that a cow is not in heat will influence the belief that the cow has mastitis
- Evidence may only be transmitted through a converging connection if a connecting variable (or descendant is observed)

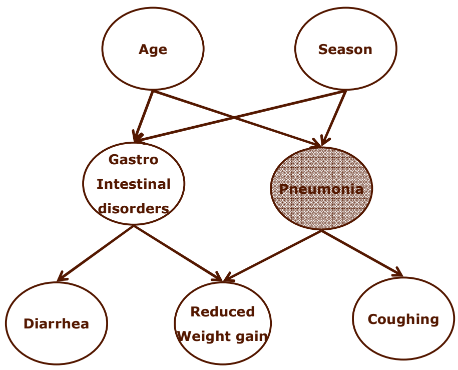
The previous example – d-separation



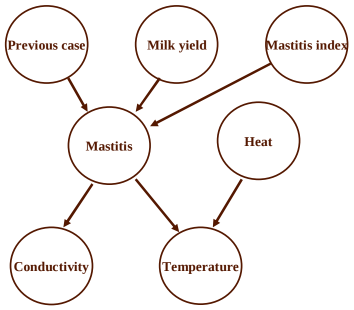
The previous example – d-separation



The previous example – d-separation



Exercise: Mastitis detection



Compilation of Bayesian networks

Cursory

Compilation:

- Create a ***moral graph***
  - Add edges between all pairs of nodes having a common child.
  - Remove all directions
- ***Triangulate the moral graph***
  - Add edges until all cycles of more than 3 nodes have a chord
- Identify the cliques of the triangulated graph and organize them into a ***junction tree***.

The software system does it automatically (and can show all intermediate stages).

Sum up

Bayesian networks

- Reasoning under uncertainty
- Graphical model with some restrictions
  - Variables and nodes form a DAG
  - All interpendencies are descibed using conditional probabilty distributions
  - Can reason against the causal direction
- Consistent combination of information from various sources
- Can estimates certainties for hypothesis variables

Next time (29th of september)

Building Bayesian networks

- Determining the graphical structure
- Determining the conditional probabilities
- Modeling tricks and tips