Graphical Models

Reference for much of the next topic is Chapter 8 of Bishop

Available on-line
http://research.microsoft.com/~cmbishop/PRML

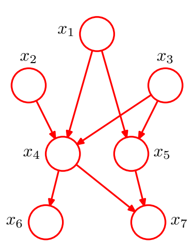
(Linked from course page).

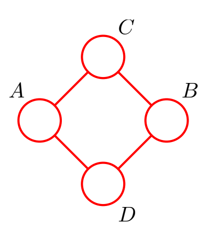
Graphical Models

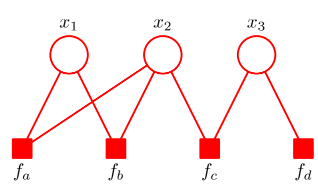
- Graphical representation of statistical models
- Nodes
  - Random variables (or groups of them)
- Edges
  - Probabilistic relationships between nodes

Graphical Models

- Various kinds
  - Directed (Bayesian networks)
  - Undirected (e.g., Markov random field)
  - Factor graphs (different representation, applicable to both)

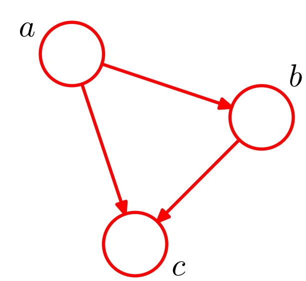






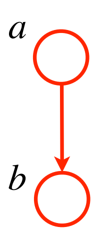
Directed Graphical Models

- Nodes represent random variables
- Edges between nodes have directed links
- No cycles

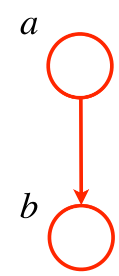


Directed Graphical Models

- Nodes represent random variables
- Edges between nodes have directed links
- No cycles
- The graph represents a **factorization** of the joint probability of all the random variables represented by the nodes.
  - An arrow from one node (a) to another one (b) means that the second node (b) is conditioned on the first (a).
  - In other words, if you have information about (a), then you have information about (b).
  - Thus the arrows tell you about information flow.



Directed Graphical Models



Here we have two nodes, a and b.

So this is a representation of the joint distribution $p(a,b)$.

In particular, it is equivalent to writing

$$p(a,b) = p(b \mid a)\ p(a)$$

Ancestral sampling version of the story:

To sample from $p(a,b)$

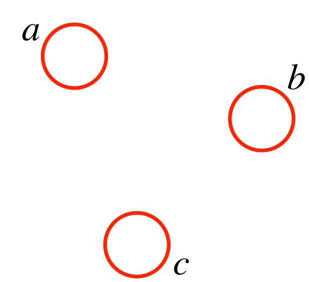
First sample $\tilde{a}$ from $p(a)$

Then sample $\tilde{b}$ from $p(b \mid \tilde{a})$

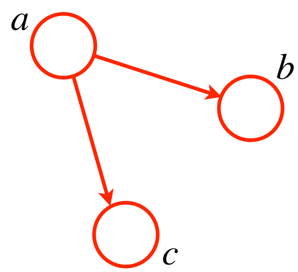
Directed Graphical Models

- A story of three random variables .... a, b, and c.
- General model is p(a,b,c) (understand this!)
- What are possible relationships of a, b, and c?
  - Independence: p(a,b,c)=p(a)p(b)p(c)
  - Some structure: e.g., p(a,b,c)=p(a)p(b|a)p(c|a)
  - Arbitrary relationship

p(a,b,c) = p(a)p(b)p(c)



p(a,b,c) = p(a)p(b|a)p(c|a)



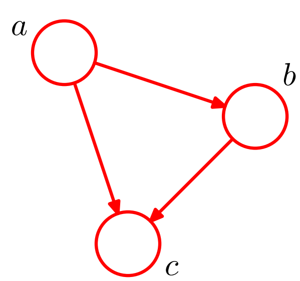
p(a,b,c) with no identified independence

$$p(a,b,c) = p(a)p(b \mid a)p(c \mid a,b)$$

$$p(a,b,c) = p(b)p(c \mid b)p(a \mid c,b)$$

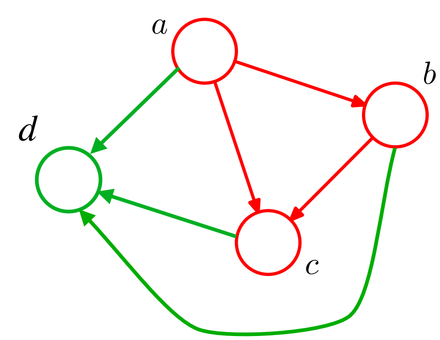
• • •

p(a,b,c) = p(a)p(b|a)p(c|a,b)



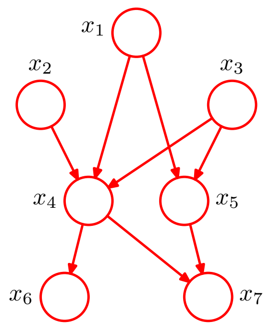
Note that the graph is fully connected

p(a,b,c,d) = p(d | a,b,c) p(a,b,c)



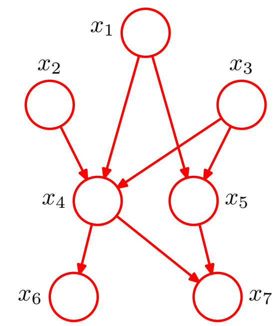
Note that the graph is fully connected

Another example (§8.2 in Bishop)



What is the algebraic form?

Another example (§8.2 in Bishop)



$$p(x_1)p(x_2)p(x_3)p(x_4 \mid x_1,x_2,x_3)p(x_5 \mid x_1,x_3)p(x_6 \mid x_4)p(x_7 \mid x_4,x_5)$$

Univariate Gaussian with known variance (§8.2 in Bishop)

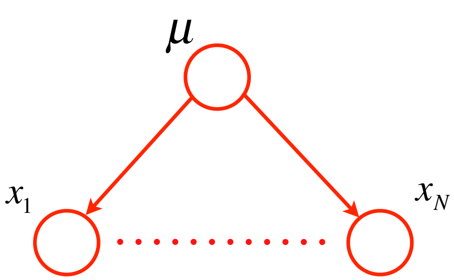
$$D = \{x_1, x_2, x_3, \ldots, x_N\}$$

$$p(D, \mu) = p(\mu)\prod_{n=1}^{N} p(x_n \mid \mu)$$

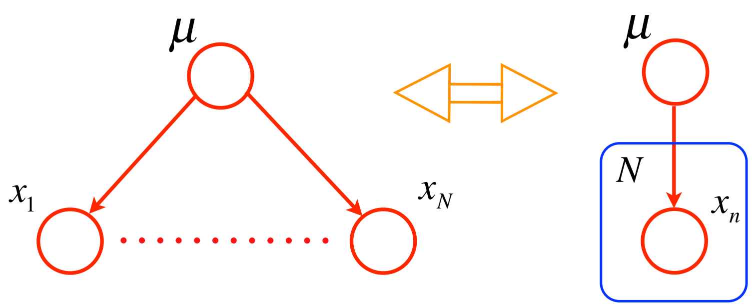
where

$$p(x_n \mid \mu) = \mathbb{N}(x_n \mid \mu; \sigma^2)$$

Univariate Gaussian with known variance (§8.2 in Bishop)



Univariate Gaussian with known variance (§8.2 in Bishop)

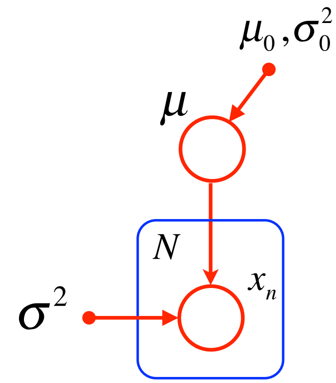


More compact notation
(plate representation)

Deterministic parameters

Our univariate Gaussian has some known parameters: the variance and the prior on the mean.

If we wish to illustrate them, we use a small filled in circle.



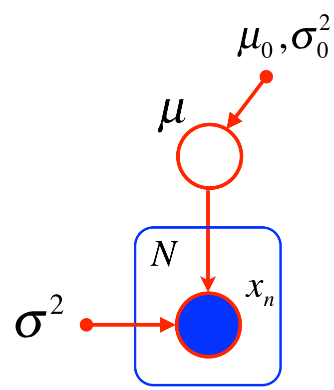
Observed variables

We indicate observed variables by shading them

Alternatively, this indicates conditioning

Observed variables

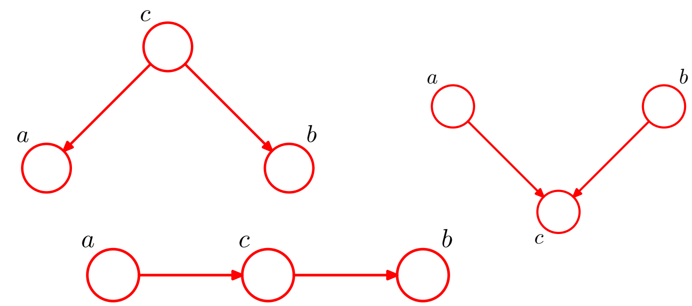
Example: Inferring the mean of the univariate



Back to three variables

What are the possible Bayes nets with three variables?

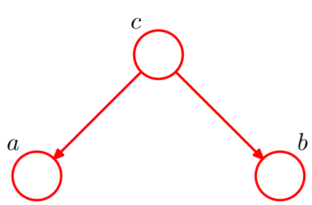
Three interesting cases



For each case, consider two questions:

1) Is $a \perp b$?
2) Is $a \perp b \mid c$? (i.e. c is observed)

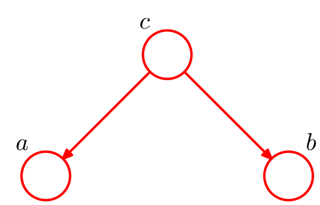
Case one



$a \not\perp b$

If you know a, that informs you about c (by Bayes) which informs you about b.

Case one

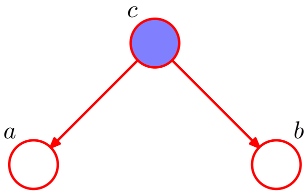


$a \not\perp b$

We can prove this with a counter example. For example, suppose *a* and *b* are coin flips, and *c* is about which of two equally likely coins are used. One is *fair* (50-50), and one is *unfair* (90,10).

Then create the contingency table for all eight possibilities. We can finish the proof by computing the marginals.

Case one where c is observed



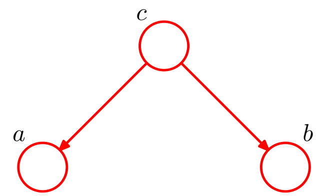
$a \perp b \mid c$

$p(a,b,c) = p(c)p(a|c)p(b|c)$ (what the graph represents in general)

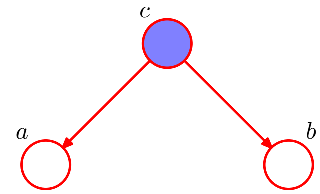
$p(a,b|c) = p(a|c)p(b|c)$ (with c observed)

This is the definition of $a \perp b|c$

Case one (tail-to-tail) summary



$a \not\perp b$

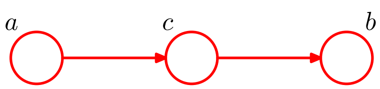


$a \perp b \mid c$

Tail-to-tail case

- With no conditioning, no independence
- With conditioning, we have independence

Case two

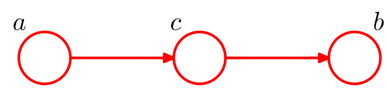


$a \not\perp b$

The graph represents $p(a,b,c) = p(a)p(c|a)p(b|c)$

If you know a, that informs you about c which informs you about b.

Case two



$a \not\perp b$

The graph represents $p(a,b,c) = p(a)p(c|a)p(b|c)$

Algebraically,

$$p(a,b) = \sum_c p(a,b,c) = p(a)\sum_c p(c|a)p(b|c)$$

If $a \perp b$ then the above is also equal to $p(a)p(b)$

$$p(a,b) = \sum_c p(a,b,c) = p(a)\sum_c p(c|a)\,p(b|c)$$

If $a \perp b$ then the above is also equal to $p(a)p(b)$

We can easily construct a counter example. Let a be a fair coin flip. Let c be a choice of unfair coin. If a is heads, then we are 90% likely to have a coin B_H biased by 90% to heads, and similarly for tails. Then,

$$p(H,H) = \frac{1}{2}\Big(p(c=B_H|a=H)\,p(b=H|c=B_H) + p(c=B_T|a=H)\,p(b=H|c=B_T)\Big)$$
$$= \frac{1}{2}(0.9*0.9+0.1*0.1)$$
$$= 0.41$$

On the other hand, by symmetry, p(a=H)=p(b=H)=$\frac{1}{2}$, and p(a=H)*p(b=H)=$\frac{1}{4}$

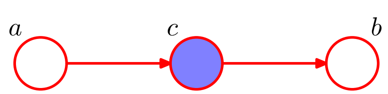
Bishop (page 374) also argues that

$$p(a,b) = \sum_c p(a,b,c) = p(a)\sum_c p(c|a)\,p(b|c) = p(a)\,p(b|a)$$

and that the RHS is not *p(a)p(b)* in general. But, when it is true,we have what we are trying to prove, so it is not (IMHO) a tight proof.

Having said that it is still worth understanding the derivation. (Hint: From the graph, *p(b|c)=p(b|a,c)*)

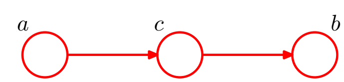
Case two where *c* is observed



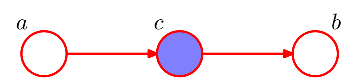
$a \perp b \mid c$

$$p(a,b\,|\,c) = \frac{p(a,b,c)}{p(c)} \quad \text{(definition)}$$
$$= \frac{p(a)\,p(c|a)\,p(b|c)}{p(c)} \quad \text{(from graph)}$$
$$= \frac{p(a)\,p(a|c)\,p(c)\,p(b|c)}{p(a)\,p(c)} \quad \text{(Bayes on } p(c|a)\text{)}$$
$$= p(a|c)\,p(b|c) \quad \text{(canceling factors)}$$

Case two (head-to-tail) summary



$a \not\perp b$

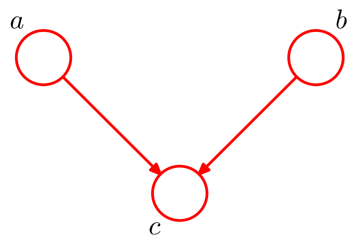


$a \perp b \mid c$

Head-to-tail case
- With no conditioning, no independence
- With conditioning, we have independence

(Same as case one)

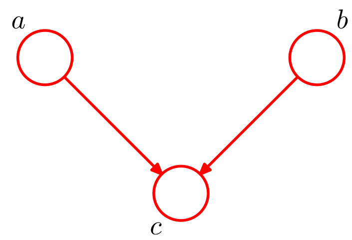
Case three



Is $a \perp b$?

Example:
c == "strange noises at night"
a == "burglar in the house"
b == "deer in the back yard"

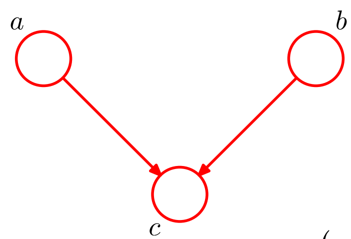
Case three



$a \perp b$

Intuitively, the arrows say that *a* is independent of *b*.

Case three

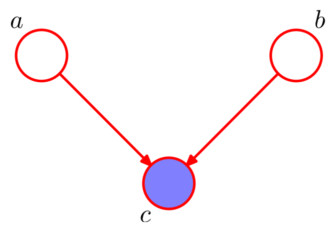


$a \perp b$

$$p(a,b,c) = p(a)\,p(b)\,p(c|a,b)$$

$$\begin{aligned} p(a,b) &= \sum_c p(a)\,p(b)\,p(c|a,b) \\ &= p(a)\,p(b)\sum_c p(c|a,b) \\ &= p(a)\,p(b) \end{aligned}$$

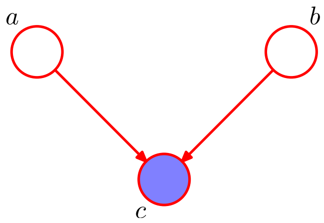
Case three with *c* observed



Is $a \perp b \mid c$?

Recall our example:
c == "strange noises at night"
a == "burglar in the house"
b == "deer in the back yard"

Case three with c observed

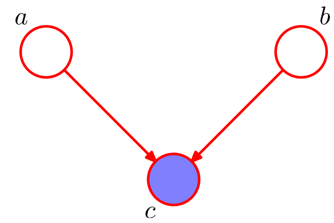


$a \not\perp b \mid c$

Intuitive explanation:

Given that we observe "strange noises", the two causes are anti-correlated (and thus not independent) due to "explaining away".

Case three with c observed

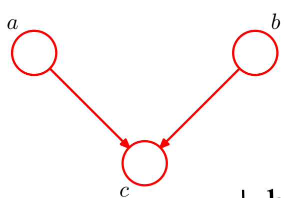


$a \not\perp b \mid c$

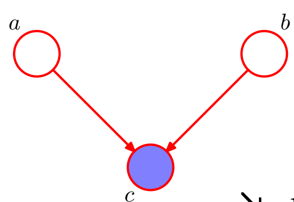
Algebraic explanation:

$$
\begin{aligned}
p(a,b|c) &= \frac{p(a,b,c)}{p(c)} \\
&= \frac{p(a)\,p(b)\,p(c|a,b)}{p(c)} \\
&= p(a)\,p(b)\frac{p(c|a,b)}{p(c)} \\
&\neq p(a)\,p(b) \qquad \text{(in general)}
\end{aligned}
$$

Case three (head-to-head) summary



$a \perp b$



$a \not\perp b \mid c$

Head-to-head case (different than the other two)

With no conditioning, we have independence

With conditioning, we do not have independence

If you are having trouble with "explaining away", please study Bishop, chapter 8, pages 378-379 (on-line).