Bayesian Nets

Introduction

- Many times, the only knowledge we have about a distribution is which variables are or are not dependent.
- Such dependencies can be represented efficiently using a Bayesian Belief Network (or Belief Net or Bayesian Net).

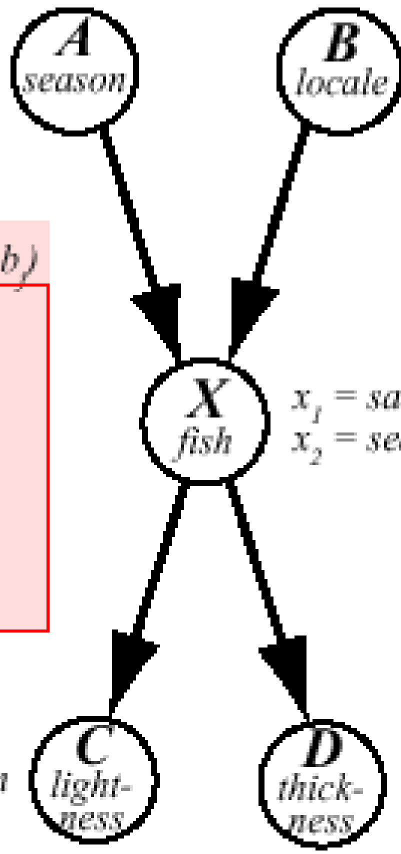
Bayesian Nets allow us to represent a joint probability density *p(x,y,z,…)* efficiently using dependency relationships.

- A belief net is usually a Directed Acyclic Graph (DAG)
- Each node represents one of the system variables.
- Each variable can assume certain states (i.e., values).

$P(a)$

$P(a_1)$	$P(a_2)$	$P(a_3)$	$P(a_4)$
0.25	0.25	0.25	0.25

a_1 = winter
a_2 = spring
a_3 = summer
a_4 = autumn



b_1 = north Atlantic
b_2 = south Atlantic

$P(b)$

$P(b_1)$	$P(b_2)$
0.6	0.4

x_1 = salmon
x_2 = sea bass

$P(x|a,b)$

	$P(x_1 \vert a_i, b_j)$	$P(x_2 \vert a_i, b_j)$
a_1, b_1	0.5	0.5
a_1, b_2	0.7	0.3
a_2, b_1	0.6	0.4
a_2, b_2	0.8	0.2
a_3, b_1	0.4	0.6
a_3, b_2	0.1	0.9
a_4, b_1	0.2	0.8
a_4, b_2	0.3	0.7

$P(c|x)$

	$P(c_1 \vert x_i)$	$P(c_2 \vert x_i)$	$P(c_3 \vert x_i)$
x_1	0.6	0.2	0.2
x_2	0.2	0.3	0.5

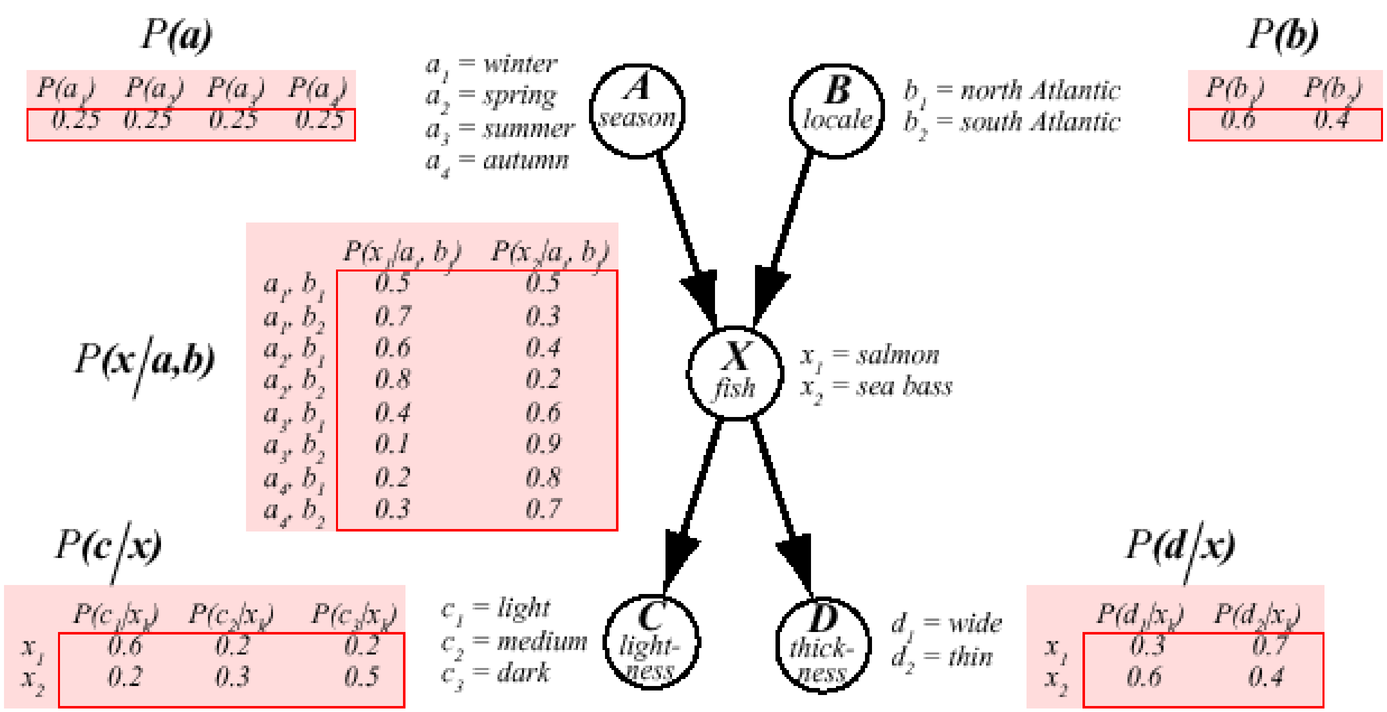
c_1 = light
c_2 = medium
c_3 = dark

d_1 = wide
d_2 = thin

$P(d|x)$

	$P(d_1 \vert x_i)$	$P(d_2 \vert x_i)$
x_1	0.3	0.7
x_2	0.6	0.4

- A link joining two nodes is directional and represents a causal influence (e.g., *X* depends on *A* or *A* influences *X*)
- Influences could be direct or indirect (e.g., *A* influences *X* directly and *A* influences *C* indirectly through *X*).



$P(a)$

$P(a_1)$	$P(a_2)$	$P(a_3)$	$P(a_4)$
0.25	0.25	0.25	0.25

$P(b)$

$P(b_1)$	$P(b_2)$
0.6	0.4

$P(x|a,b)$

	$P(x_1|a_i, b_j)$	$P(x_2|a_i, b_j)$
a_1, b_1	0.5	0.5
a_1, b_2	0.7	0.3
a_2, b_1	0.6	0.4
a_2, b_2	0.8	0.2
a_3, b_1	0.4	0.6
a_3, b_2	0.1	0.9
a_4, b_1	0.2	0.8
a_4, b_2	0.3	0.7

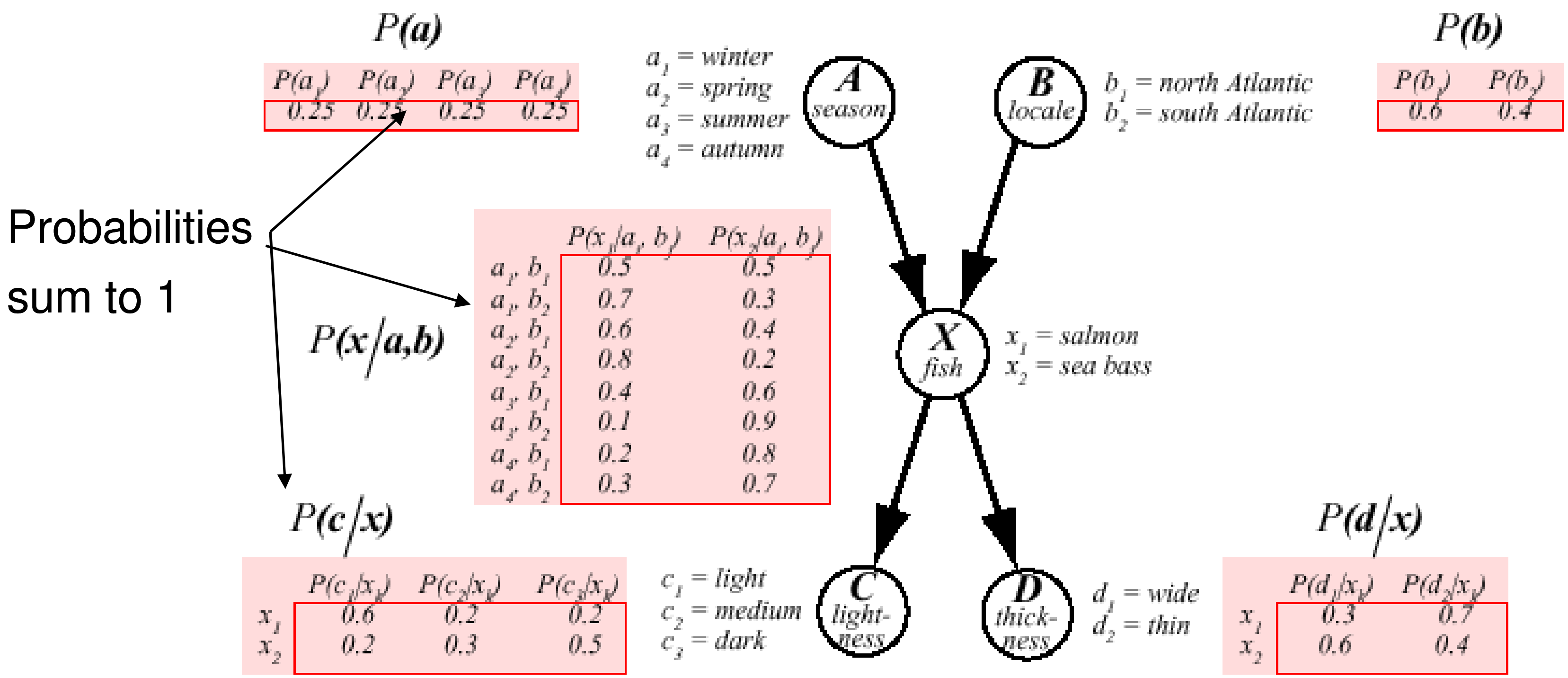
$P(c|x)$

	$P(c_1|x_k)$	$P(c_2|x_k)$	$P(c_3|x_k)$
x_1	0.6	0.2	0.2
x_2	0.2	0.3	0.5

$P(d|x)$

	$P(d_1|x_k)$	$P(d_2|x_k)$
x_1	0.3	0.7
x_2	0.6	0.4

- Each variable is associated with a set of weights which represent prior or conditional probabilities (discrete or continuous) .



- Using the chain rule, the joint probability of a set of variables $x_1, x_2, \ldots, x_n$ is give as:

$$p(x_1, x_2, \ldots, x_n) = p(x_1 / x_2, \ldots, x_n)\, p(x_2 / x_3, \ldots, x_n) \ldots p(x_{n-1} / x_n)\, p(x_n)$$

- The conditional independence relationships encoded in the Bayesian network state that a node x_i is conditionally independent of its ancestors given its parents π_i :

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i / \pi_i) \Rightarrow$$ much simpler!

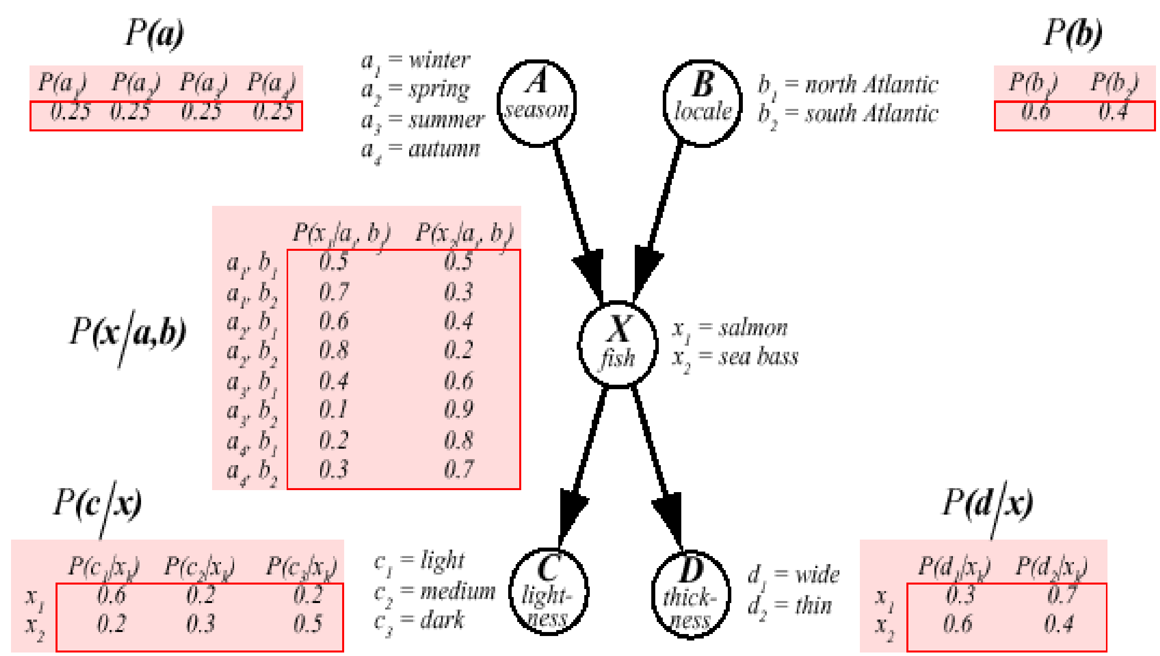
- We can compute the probability of any configuration of variables in the joint density distribution, e.g.:

(probability of catching a medium lightness, thin sea-bass from the North Atlantic in summer)

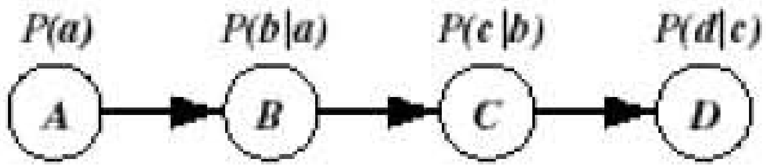
$P(a_3, b_1, x_2, c_3, d_2)$

$= P(a_3)P(b_1)P(x_2 / a_3,b_1)P(c_3 / x_2)P(d_2 / x_2)$

$= 0.25 \times 0.6 \times 0.4 \times 0.5 \times 0.4 = 0.012$



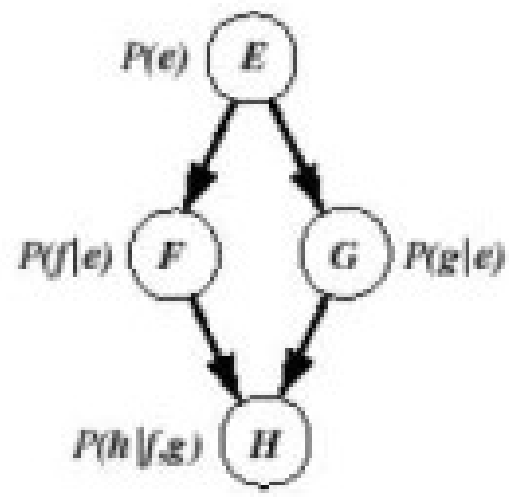
- e.g., determine the probability at D



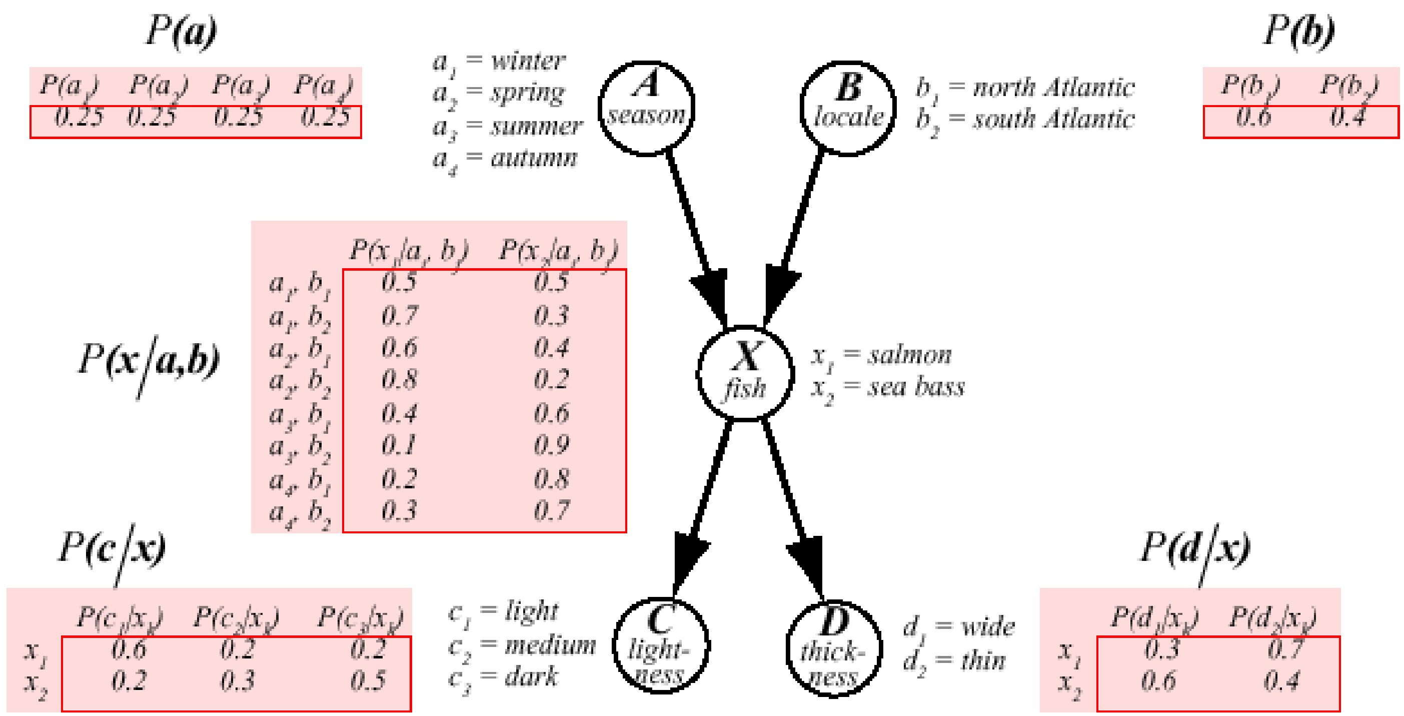
$$\begin{aligned} P(\mathrm{d}) &= \sum_{\mathrm{a,b,c}} P(\mathrm{a,b,c,d}) \\ &= \sum_{\mathrm{a,b,c}} P(\mathrm{a})P(\mathrm{b}|\mathrm{a})P(\mathrm{c}|\mathrm{b})P(\mathrm{d}|\mathrm{c}) \\ &= \underbrace{\sum_{\mathrm{c}} P(\mathrm{d}|\mathrm{c}) \underbrace{\sum_{\mathrm{b}} P(\mathrm{c}|\mathrm{b}) \underbrace{\sum_{\mathrm{a}} P(\mathrm{b}|\mathrm{a})P(\mathrm{a})}_{P(\mathrm{b})}}_{P(\mathrm{c})}}_{P(\mathrm{d})} \end{aligned}$$

- e.g., determine the probability at H:

$$
\begin{aligned}
P(\mathrm{h}) &= \sum_{\mathrm{e,f,g}} P(\mathrm{e,f,g,h}) \\
&= \sum_{\mathrm{e,f,g}} P(\mathrm{e})P(\mathrm{f}|\mathrm{e})P(\mathrm{g}|\mathrm{e})P(\mathrm{h}|\mathrm{f,g}) \\
& \sum_{\mathrm{f,g}} P(\mathrm{h}|\mathrm{f,g}) \sum_{\mathrm{e}} P(\mathrm{e})P(\mathrm{f}|\mathrm{e})P(\mathrm{g}|\mathrm{e})
\end{aligned}
$$



- Classify a fish given that the fish is light (c_1) and was caught in south Atlantic (b_2) -- no evidence about what time of the year the fish was caught nor its thickness.



$$
\begin{aligned}
P(x_1|c_1,b_2) &= \frac{P(x_1,c_1,b_2)}{P(c_1,b_2)} \\
&= \alpha \sum_{\mathrm{a,d}} P(x_1,\mathbf{a},b_2,c_1,\mathbf{d}) \\
&= \alpha \sum_{\mathrm{a,d}} P(\mathbf{a})P(b_2)P(x_1|\mathbf{a},b_2)P(c_1|x_1)P(\mathbf{d}|x_1) \\
&= \alpha P(b_2)P(c_1|x_1) \\
&\quad \times \left[\sum_{\mathrm{a}} P(\mathbf{a})P(x_1|\mathbf{a},b_2)\right]\left[\sum_{\mathrm{d}} P(\mathbf{d}|x_1)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \alpha P(b_2)P(c_1|x_1) \\
&\quad \times [P(a_1)P(x_1|a_1,b_2) + P(a_2)P(x_1|a_2,b_2) + P(a_3)P(x_1|a_3,b_2) + P(a_4)P(x_1|a_4,b_2)] \\
&\quad \times \underbrace{[P(d_1|x_1) + P(d_2|x_1)]}_{=1} \\
&= \alpha(0.4)(0.6)\,[(0.25)(0.7) + (0.25)(0.8) + (0.25)(0.1) + (0.25)(0.3)]\;1.0 \\
&= \alpha\;0.114.
\end{aligned}
$$

- Similarly,

$$P(x_2 / c_1, b_2) = \alpha\ 0.066$$

- Normalize probabilities (not needed necessarily):

$$P(x_1 / c_1, b_2) + P(x_2 / c_1, b_2) = 1 \quad (\alpha = 1/0.18)$$

$$P(x_1 / c_1, b_2) = 0.73$$

$$P(x_2 / c_1, b_2) = 0.27$$