1. Bayesian Network

1. Graphic Model & it's expression

For N random variables, if we want to describle its whole behavior, we need to store the joint distribution

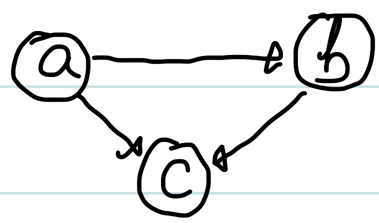
$$p(x_1, x_2, \cdots x_N)$$

Even if x_i are binary, we still need at least 2^N-1 values.
However, if we think of the case where $x_1, \cdots x_N$ are mutually independent, then

$$p(x_1, \cdots x_N) = \prod_{i=1}^{N} p(x_i)$$

we only need to store N values. That is a huge improvement.
So, the basic idea is to use the independency to simplify the joint distribution.

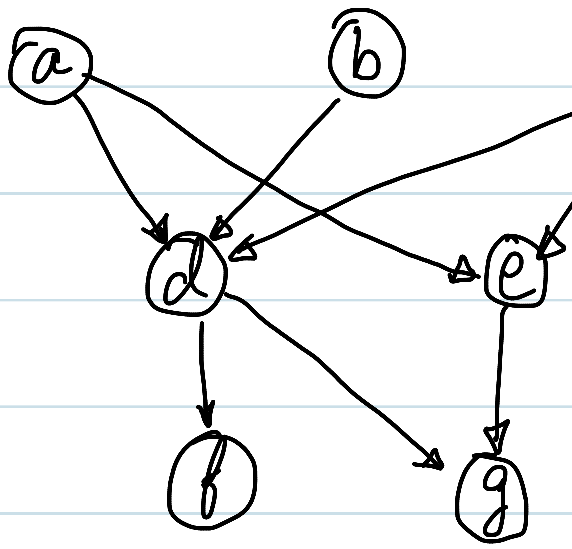
If $p(a, b, c) = p(c|a,b)\, p(b|a)\, p(a)$, No independency
which is the chain rule. we can draw the graph



We find that the network is fully connected, which means no independence.

If $p(a,b,c,d,e,f,g)$

$$= p(a)\,p(b)\,p(c)\,p(d|a,b,c)\,p(e|a,c)\,p(f|d)\,p(g|d,e)$$

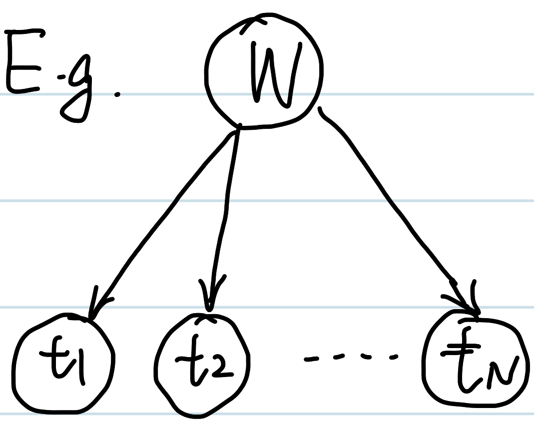


Then, we have the following formula

$$P(\underset{\sim}{X}) = \prod_{i=1}^{n} p(x_i \mid px_i)$$

px_i refers to x_i's parents node.

E.g.

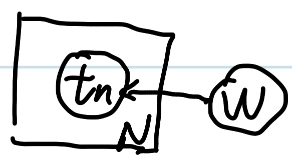


This is "polynomial regression".

$$p(\underline{t}, w) = p(w)\cdot \prod_{i=1}^{n} p(t_i|w)$$

$t_1 \ldots t_N$ are data, w is the polynomial coeff.

Usually, we can use the following figure



Eg. Given the following expression, draw the corresponding graph. (linear regression)

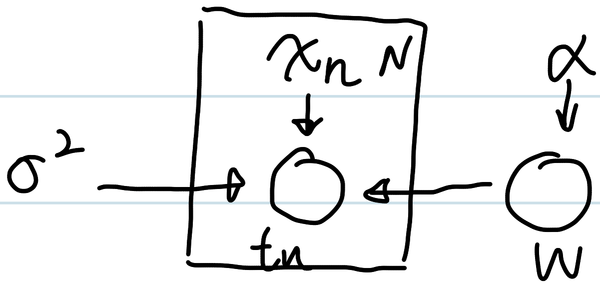
$$p(t, w \mid x, \alpha, \sigma^2) = p(w \mid \alpha) \prod_{n=1}^{N} p(t_n \mid w, x_n, \sigma^2)$$

From the expression, we know

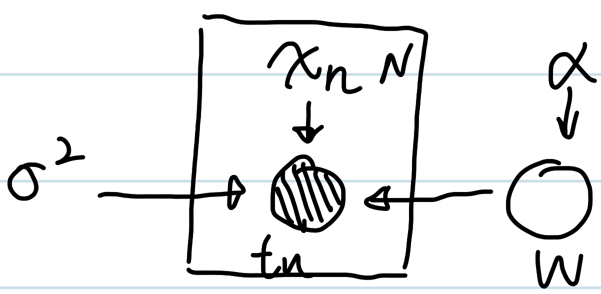
$$\alpha \to w$$

$$w, x_n, \sigma^2 \to t_n$$

$\therefore$ we have



We usually use shadowed circle to represent "observed" variables. For most cases, in the training set we know t_n, x_n, we want to know w, so the graph turns to



A stright forward Bayes theorem application is

$$p(\underline{w} \mid \underline{t}) \propto p(\underline{w}) \prod_{n=1}^{N} p(t_n \mid w)$$

2. Conditional Independence.

If we have

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

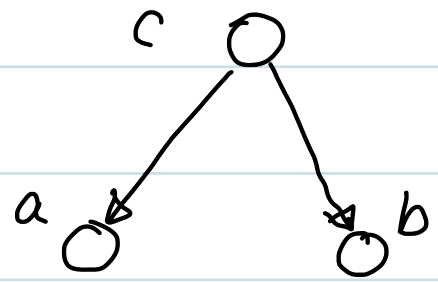
we say that $a \perp b | c$

Note that, conditional independence $\not\Leftrightarrow$ independence.

Rewrite it,

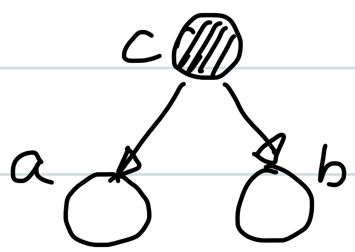
$$p(a,b|c) = p(a|b,c) \cdot p(b|c)$$
$$= p(a|c)\, p(b|c)$$

which leads to the following graph



Three basic example graphs of three R.Vs.

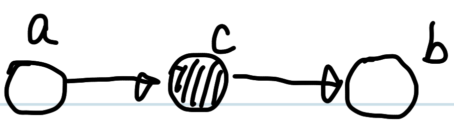
a)



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

or $a \perp b | c$

b)

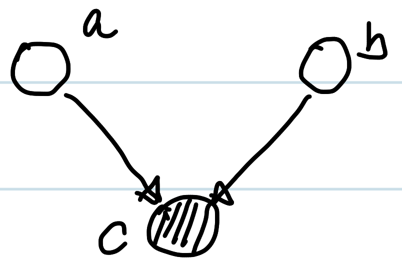


$\because p(a,b,c) = p(a)\, p(c|a)\, p(b|c)$

$$\therefore p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)}{p(c)} p(b|c)$$
$$= p(a|c)\, p(b|c)$$

also $a \perp b | c$

c)



$$P(a,b,c) = P(a)P(b)P(c|a,b)$$

$$P(a,b|c) = \frac{P(a)P(b)P(c|a,b)}{P(c)} \neq P(a|c)P(b|c)$$

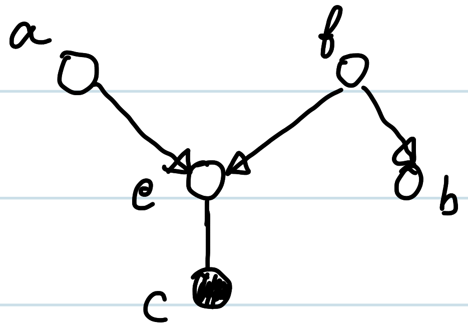
So $a \not\perp b \mid c$

Now let's generalize the conclusion of above three examples to larger graph.

- From example a), if the parent node is observed, it's descendants are conditional independent
- From example b), it's a Markov Chain. For a sequential dependent relationship, the parent node is given, then it's descentant & grand parent node are conditional independent.
- From example c), if the descendant is given, its parents are <u>not</u> conditional independent.

Usually, people are fund of using "block" to describe the ending of dependance. for example b), "dependence is block by node C".

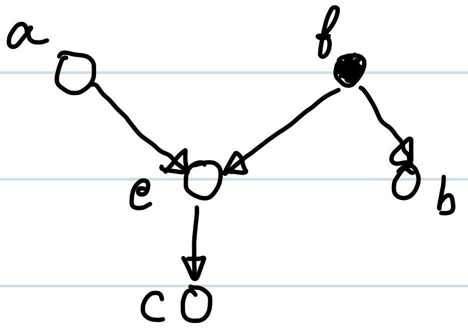
Some more examples.



the path from a → b is not blocked.

First, it's not blocked by f, because f is not observed

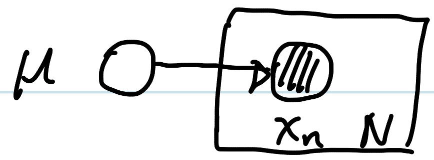
Second, it's not blocked by c, because c is given, or say c is in the conditioning set.



the path from a → b is blocked by f.

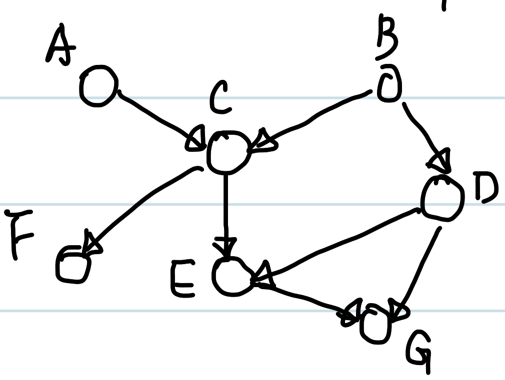
This is called "d-separation" theorem.

A special case is iid experiment. It's true that there is path



from x_i to $x_{j\neq i}$, but it's blocked by x_i. So, it's conditional independent.

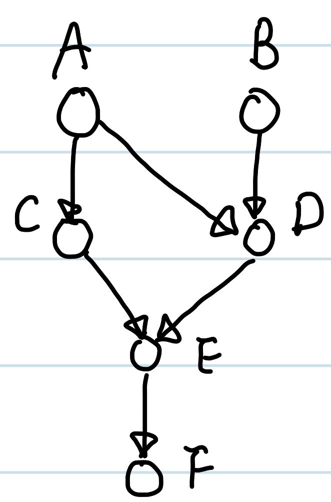
Last two exampls.



$D \perp F \mid C$

$A \perp G \mid D, E$

$A \not\perp B \mid E$



$A \perp B \mid C$

$B \not\perp E \mid D$

$C \not\perp D \mid F$