DS-GA 3001.001/002
Probabilistic Time Series Analysis
Lab 1

Caroline Haimerl

NYU

Logistics

- Time & Location
  - Wednesday 3:30pm-4:20pm, room 110
- TA Office Hour
  - Thursday 11am, room TBA
- Lab HW Submission
  - submit your lab work Wednesday by 3pm

https://github.com/savinteachingorg/pTSAfall2019

Bayes recap

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

$$p(X = x_i)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \dots, M$ and $j = 1, \dots, L$

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_i)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \dots, M$ and $j = 1, \dots, L$

the **sum** rule of probability (law of total probability)

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_i)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

the **sum** rule of probability (law of total probability)

$$\underbrace{p(X = x_i)}_{\text{marginal probability}} = \sum_{j=1}^{L} p(X = x_i, Y = y_i)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

the **sum** rule of probability (law of total probability)

$$\underbrace{p(X = x_i)}_{\text{marginal probability}} = \sum_{j=1}^{L} \underbrace{p(X = x_i, Y = y_j)}_{\text{joint probability}}$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

$$p\left(X = x_i, Y = y_j\right) = p\left(X = x_i \middle| Y = y_j\right) p(Y = y_j)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

$$p(X = x_i, Y = y_j) = \underbrace{p(X = x_i | Y = y_j)}_{\text{conditional probability}} p(Y = y_j)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \dots, M$ and $j = 1, \dots, L$

the **product** rule of probability

$$p\left(X = x_i, Y = y_j\right) = \underbrace{p\left(X = x_i \middle| Y = y_j\right)}_{\text{conditional probability}} p(Y = y_j)$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

Bayes‘ theorem

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \ldots, M$ and $j = 1, \ldots, L$

Bayes' theorem

$$p\left(Y = y_j \middle| X = x_i\right) = \frac{p\left(X = x_i \middle| Y = y_j\right) p\left(Y = y_j\right)}{p(X = x_i)}$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \dots, M$ and $j = 1, \dots, L$

Bayes‘ theorem ... to reverse conditional probabilities

$$p\left(X = x_i \middle| Y = y_j\right) = \frac{p\left(Y = y_j \middle| X = x_i\right) p(X = x_i)}{p(Y = y_j)}$$

Bayes recap

given random variables X, Y that can take values $\{x_i\}$ and $\{y_j\}$ respectively, where $i = 1, \dots, M$ and $j = 1, \dots, L$

Bayes' theorem ... to reverse conditional probabilities

$$p\left(X = x_i \middle| Y = y_j\right) = p(X|Y)$$

Bayes recap

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

Bayes recap

$$p(X|Y) = \frac{p(Y|X)\color{red}{p(X)}}{p(Y)}$$

prior probability

... available before observation

Bayes recap

likelihood

... of data

prior probability

... available before observation

$$p(X|Y) = \frac{\color{red}{p(Y|X)}p(X)}{p(Y)}$$

Bayes recap

likelihood

... of data

prior probability

... available before observation

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

normalization constant

Bayes recap

likelihood
... of data

posterior probability
... after observation

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

prior probability
... available before observation

normalization constant

Example 2.3. Jo has a test for a nasty disease. We denote Jo's state of health by the variable a and the test result by b.

$$\begin{array}{ll} a = 1 & \text{Jo has the disease} \\ a = 0 & \text{Jo does not have the disease.} \end{array} \tag{2.12}$$

The result of the test is either 'positive' ($b = 1$) or 'negative' ($b = 0$); the test is 95% reliable: in 95% of cases of people who really have the disease, a positive result is returned, and in 95% of cases of people who do not have the disease, a negative result is obtained. The final piece of background information is that 1% of people of Jo's age and background have the disease.

OK – Jo has the test, and the result is positive. What is the probability that Jo has the disease?

from MacKay (1995)

Solution. We write down all the provided probabilities. The test reliability specifies the conditional probability of b given a:

$$P(b=1 \mid a=1) = 0.95 \qquad P(b=1 \mid a=0) = 0.05$$
$$P(b=0 \mid a=1) = 0.05 \qquad P(b=0 \mid a=0) = 0.95; \tag{2.13}$$

and the disease prevalence tells us about the marginal probability of a:

$$P(a=1) = 0.01 \qquad P(a=0) = 0.99. \tag{2.14}$$

From the marginal $P(a)$ and the conditional probability $P(b \mid a)$ we can deduce the joint probability $P(a, b) = P(a)P(b \mid a)$ and any other probabilities we are interested in. For example, by the sum rule, the marginal probability of $b=1$ – the probability of getting a positive result – is

$$P(b=1) = P(b=1 \mid a=1)P(a=1) + P(b=1 \mid a=0)P(a=0). \tag{2.15}$$

Jo has received a positive result $b=1$ and is interested in how plausible it is that she has the disease (i.e., that $a=1$). The man in the street might be duped by the statement 'the test is 95% reliable, so Jo's positive result implies that there is a 95% chance that Jo has the disease', but this is incorrect. The correct solution to an inference problem is found using Bayes' theorem.

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1 \mid a=1)P(a=1) + P(b=1 \mid a=0)P(a=0)} \tag{2.16}$$
$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} \tag{2.17}$$
$$= 0.16. \tag{2.18}$$

So in spite of the positive result, the probability that Jo has the disease is only 16%. □

from MacKay (1995)

Bayes recap

$$p(X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) =$$

Bayes recap

$$p(X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) =$$

$$p(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_0 = x_0) * p(X_{t-1} = x_{t-1} | X_{t-2} = x_{t-2}, \dots X_0 = x_0) * \dots * p(X_0 = x_0)$$

Bayes recap

$$p(X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) =$$

$$\prod_k p(X_k = x_k | X_{k-1} = x_{k-1}, \dots, X_0 = x_0)$$

Bayes recap

the **chain rule**

$$p(X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) =$$

$$\prod_k p(X_k = x_k | X_{k-1} = x_{k-1}, \dots, X_0 = x_0)$$

graphical models

graphical models

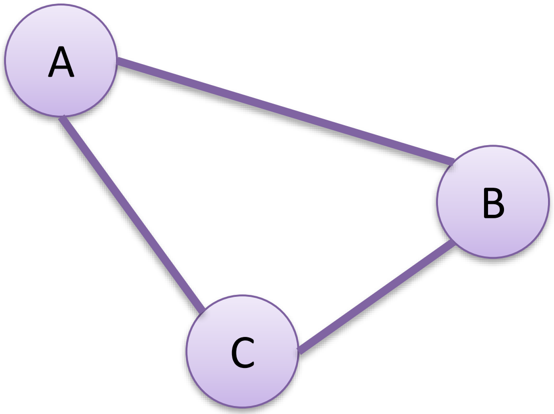
joint distributions of three variables

$$p(A, B, C) =$$

graphical models

joint distributions of three variables

$$p(A, B, C) =$$

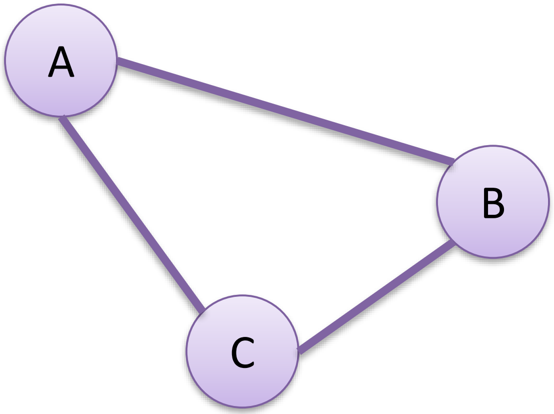


graphical models

joint distributions of three variables

$$p(A, B, C) =$$

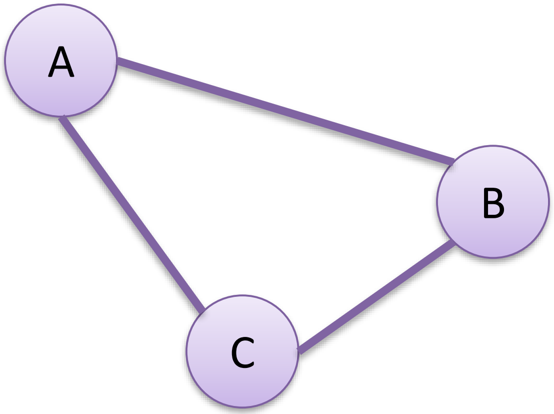
Markov Random Fields
= undirected graphical models



graphical models

joint distributions of three variables

$$p(A, B, C) = p(C|A, B)p(B|A)p(A)$$

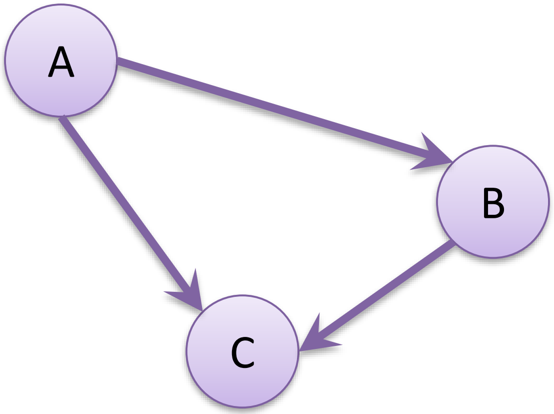


graphical models

joint distributions of three variables

$$p(A, B, C) = p(C|A, B)p(B|A)p(A)$$

Bayesian networks
= directed graphical models

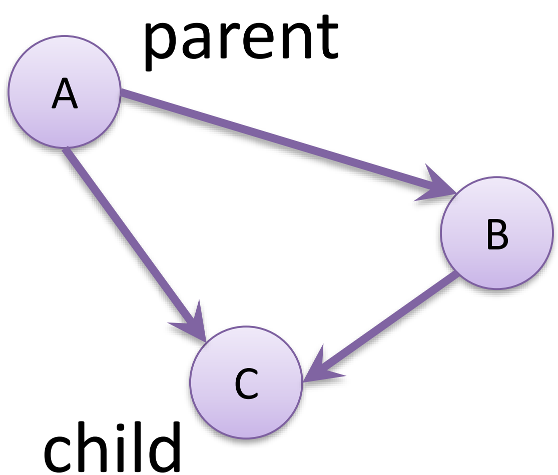


graphical models

joint distributions of three variables

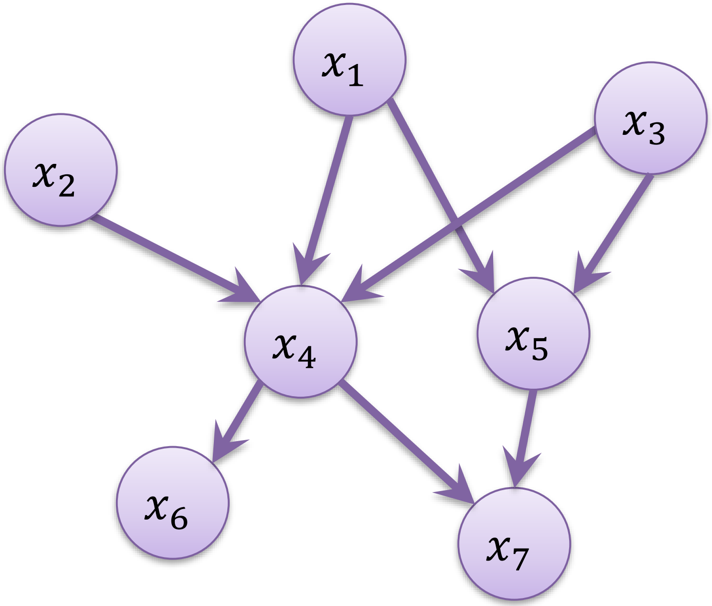
$$p(A, B, C) = p(C|A, B)p(B|A)p(A)$$

Bayesian networks
= directed graphical models



graphical models

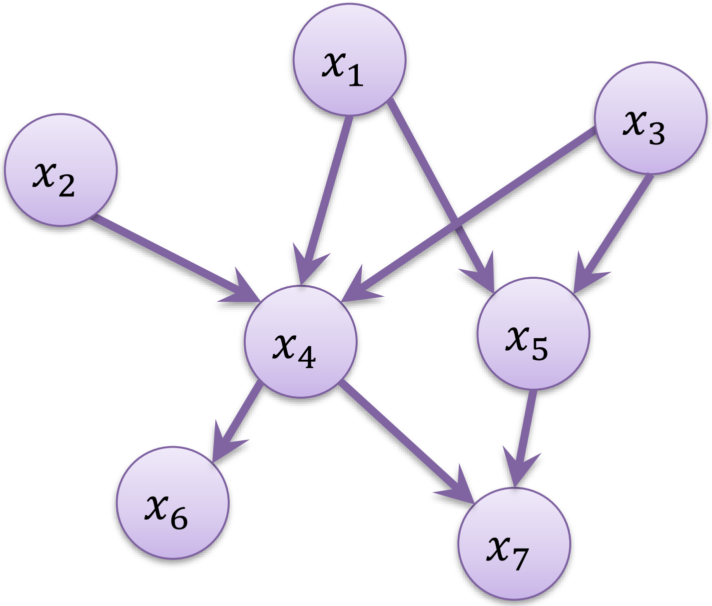
Bayesian networks



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

graphical models

Bayesian networks

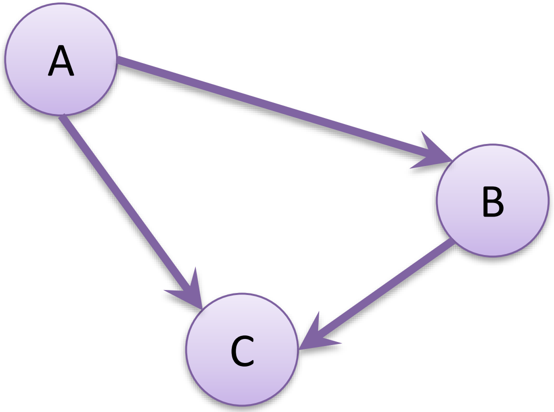


$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7) =$$
$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

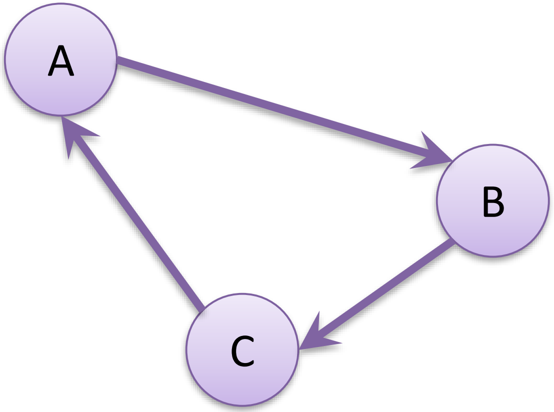
graphical models

Bayesian networks

directed acyclic graphs (DAG)



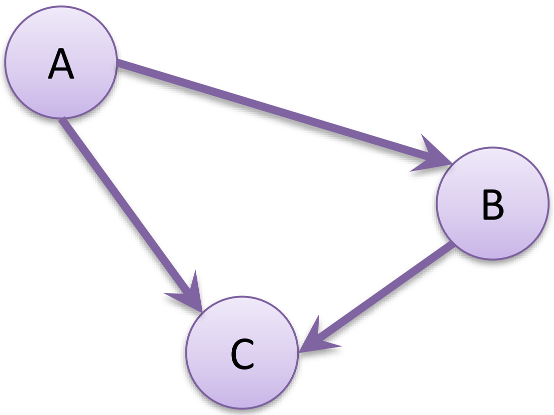
directed cyclic graphs



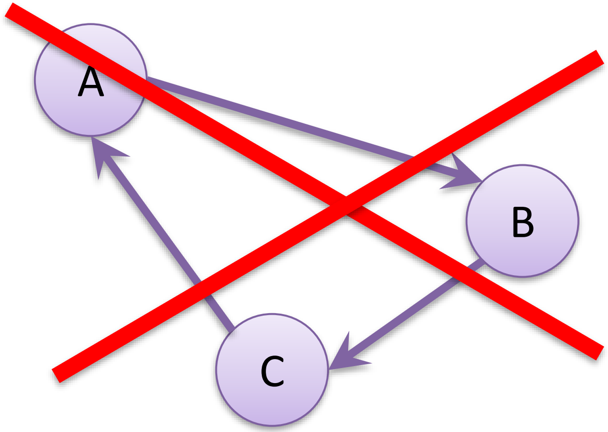
graphical models

Bayesian networks

directed acyclic graphs (DAG)



directed cyclic graphs



independence

A and B are **independent** if $p(A,B) = p(A)p(B)$

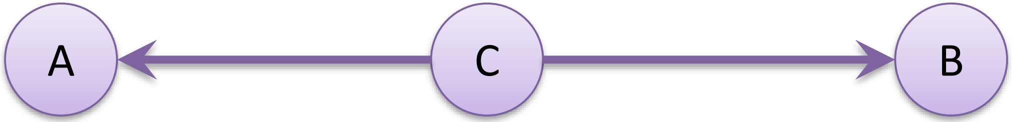
conditional independence

A and B are **conditionally independent** given C if

$$p(A, B|C) = p(A|B, C)p(B|C) = p(A|C)p(B|C)$$

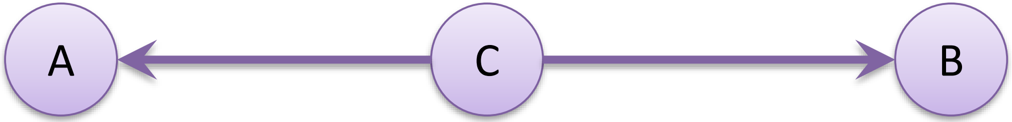
conditional independence

conditional independence



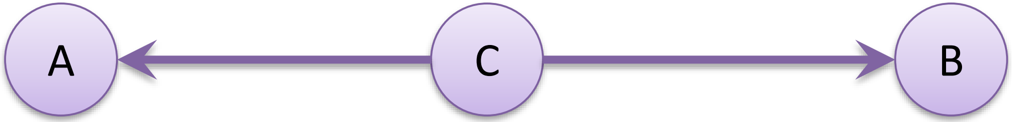
$$p(A, B, C) =$$

conditional independence



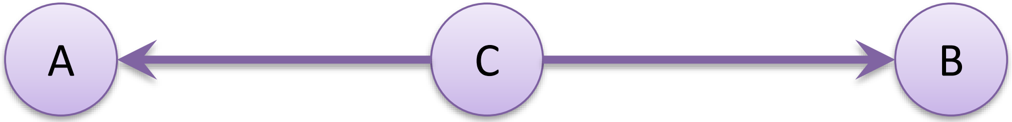
$$p(A, B, C) = p(A|B, C)p(B|C)p(C) =$$

conditional independence



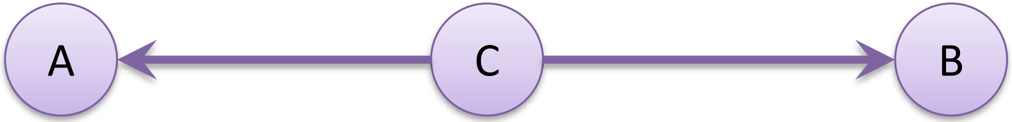
$$p(A,B,C) = p(A|B,C)p(B|C)p(C) = p(A|C)p(B|C)p(C)$$

conditional independence



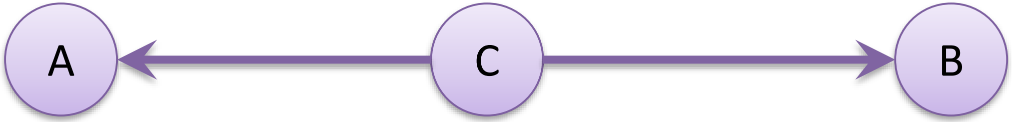
$p(A, B) =$ $\quad ?\, p(A)\, p(B)$

conditional independence



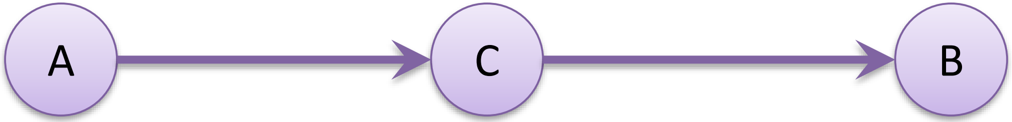
$$p(A,B) = \sum_C p(A|C)p(B|C)p(C) \quad ?\, p(A)\, p(B)$$

conditional independence



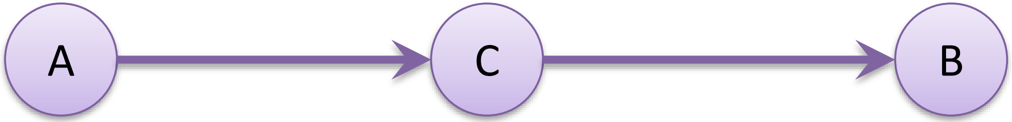
$$p(A,B) = \sum_C p(A|C)p(B|C)p(C) \quad \neq p(A)\,p(B)$$

conditional independence



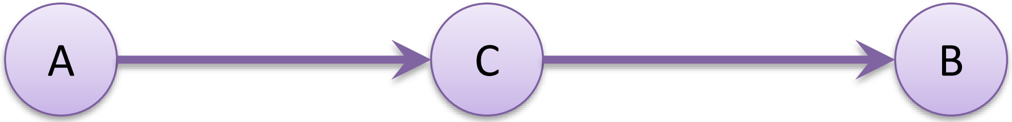
$$p(A, B, C) =$$

conditional independence



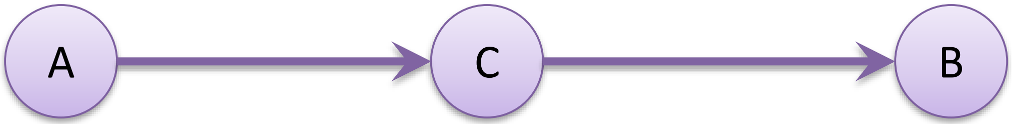
$$p(A,B,C) = p(B|A,C)p(C|A)p(A) = p(B|C)p(C|A)p(A)$$

conditional independence



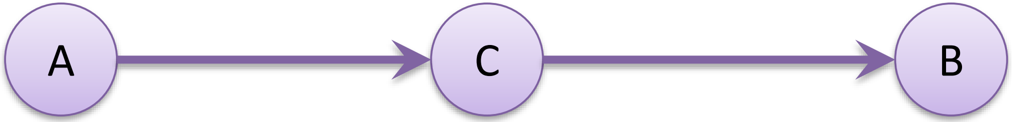
$p(A, B) =$ $\quad ?\, p(A)\, p(B)$

conditional independence



$$p(A,B) = p(A)\sum_{C} p(C|A)p(B|C) \neq p(A)\,p(B)$$

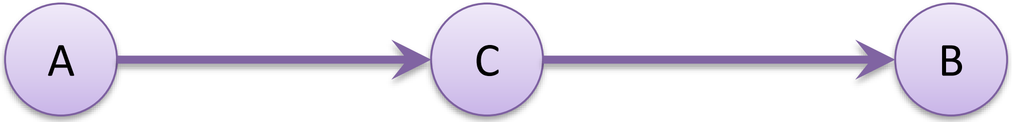
conditional independence



$$p(A, B|C) =$$

$$? \, p(A|C)p(B|C)$$

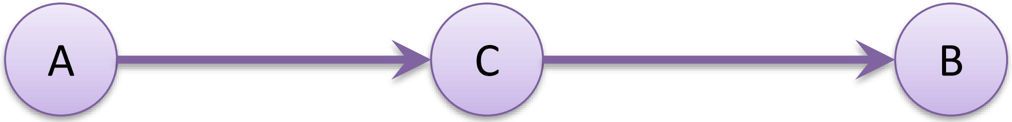
conditional independence



$$p(A,B|C) = \frac{p(A,B,C)}{p(C)} =$$

$$? \, p(A|C)p(B|C)$$

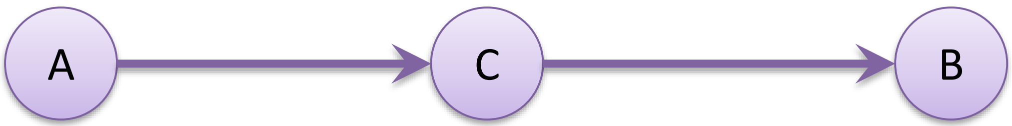
conditional independence



$$p(A,B|C) = \frac{p(A,B,C)}{p(C)} = \frac{p(A)p(C|A)p(B|C)}{p(C)}$$

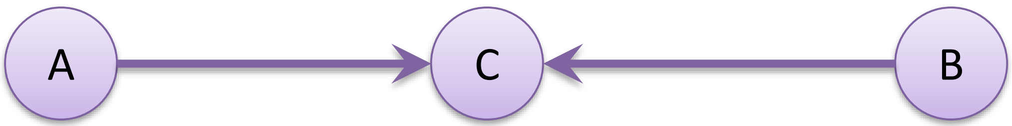
$$? \, p(A|C)p(B|C)$$

conditional independence



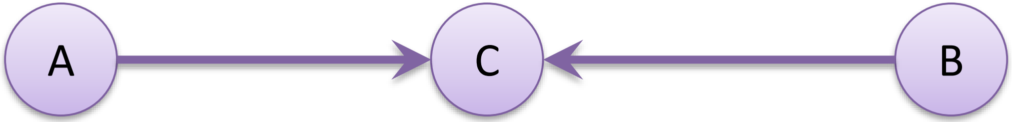
$$p(A,B|C)=\frac{p(A,B,C)}{p(C)}=\frac{p(A)p(C|A)p(B|C)}{p(C)}$$
$$=p(A|C)p(B|C)$$

conditional independence



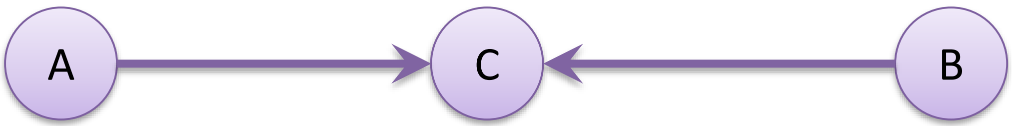
$$p(A, B, C) =$$

conditional independence



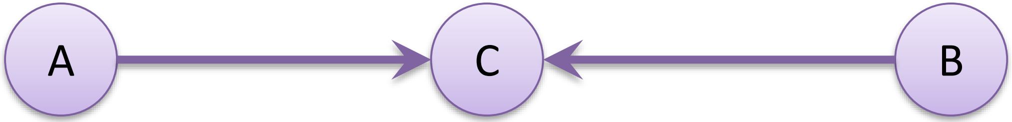
$$p(A, B, C) = p(A)p(B)p(C|A, B)$$

conditional independence



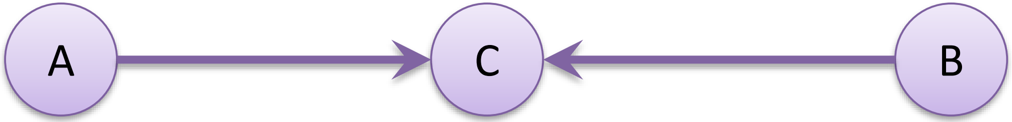
$p(A, B)$ $? \; p(A)p(B)$

conditional independence



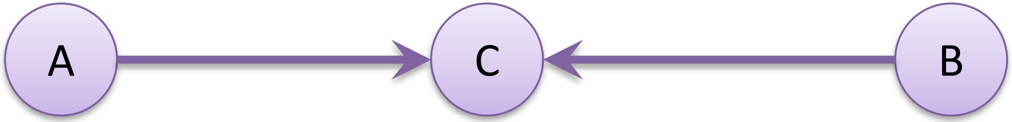
$$p(A,B) = p(A)p(B)\sum_{C} p(C|A,B) \; ? \; p(A)p(B)$$

conditional independence



$$p(A,B) = p(A)p(B)\sum_{C} p(C|A,B) = p(A)p(B)$$

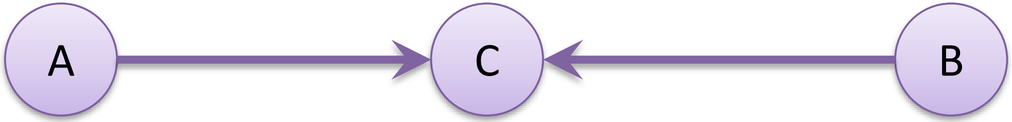
conditional independence



$$p(A, B|C) =$$

$$? \, p(A|C)p(B|C)$$

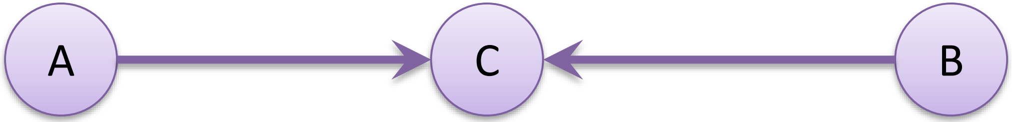
conditional independence



$$p(A,B|C) = \frac{p(A,B,C)}{p(C)} = \frac{p(A)p(B)p(C|A,B)}{p(C)}$$

$$?\, p(A|C)p(B|C)$$

conditional independence

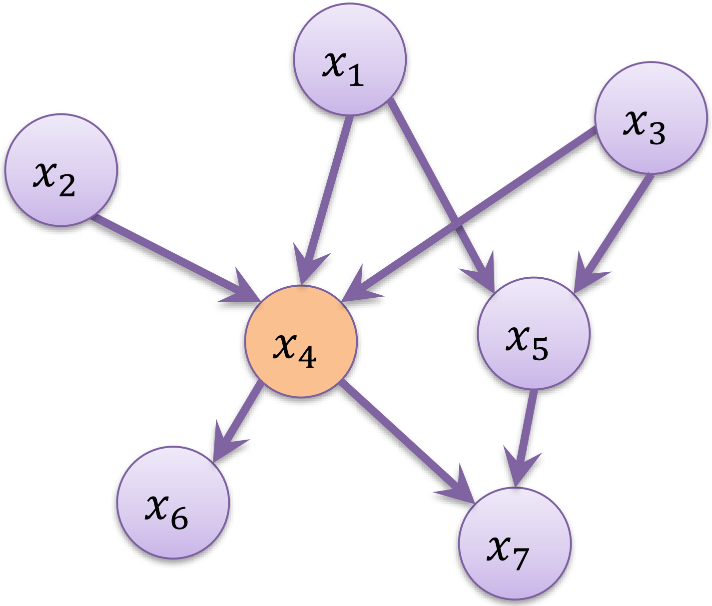


$$p(A,B|C) = \frac{p(A,B,C)}{p(C)} = \frac{p(A)p(B)p(C|A,B)}{p(C)}$$

$$\neq p(A|C)p(B|C)$$

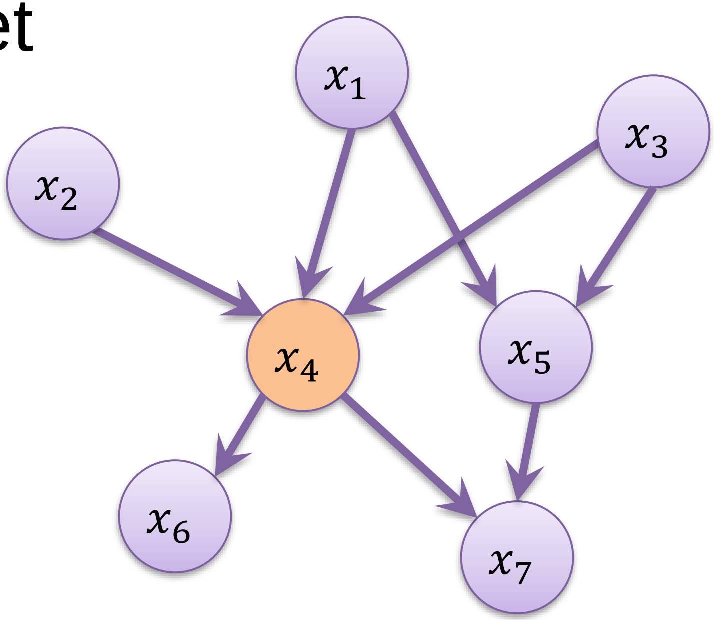
conditional independence

d-separated



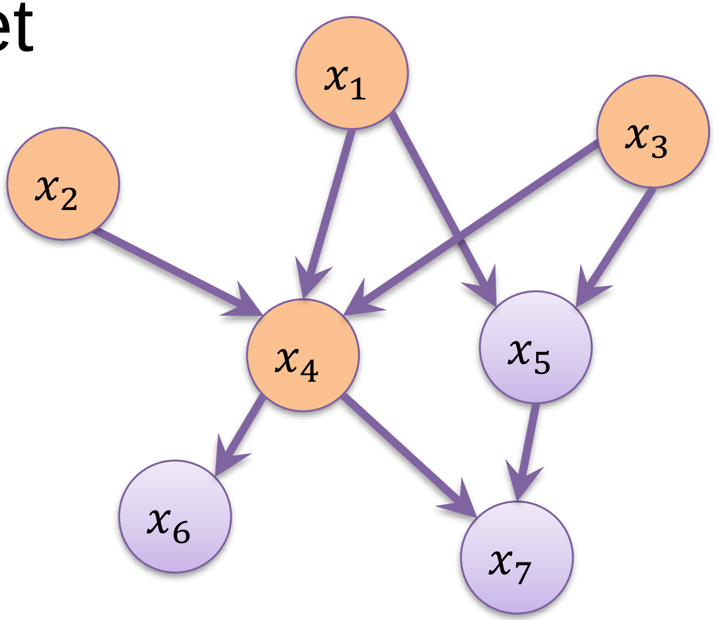
conditional independence

Markov blanket



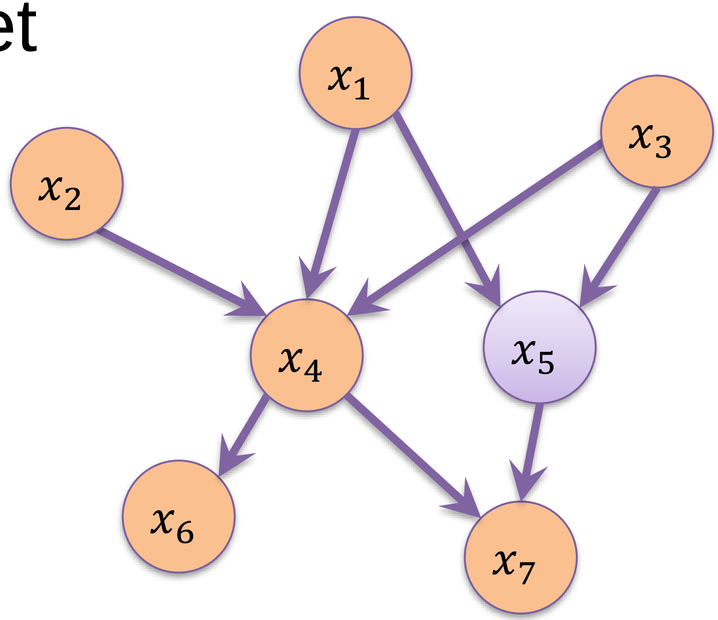
conditional independence

Markov blanket



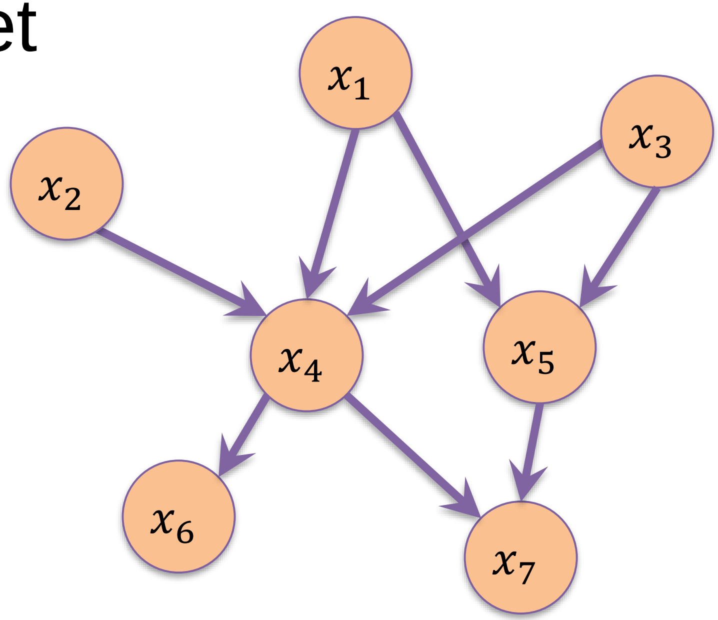
conditional independence

Markov blanket



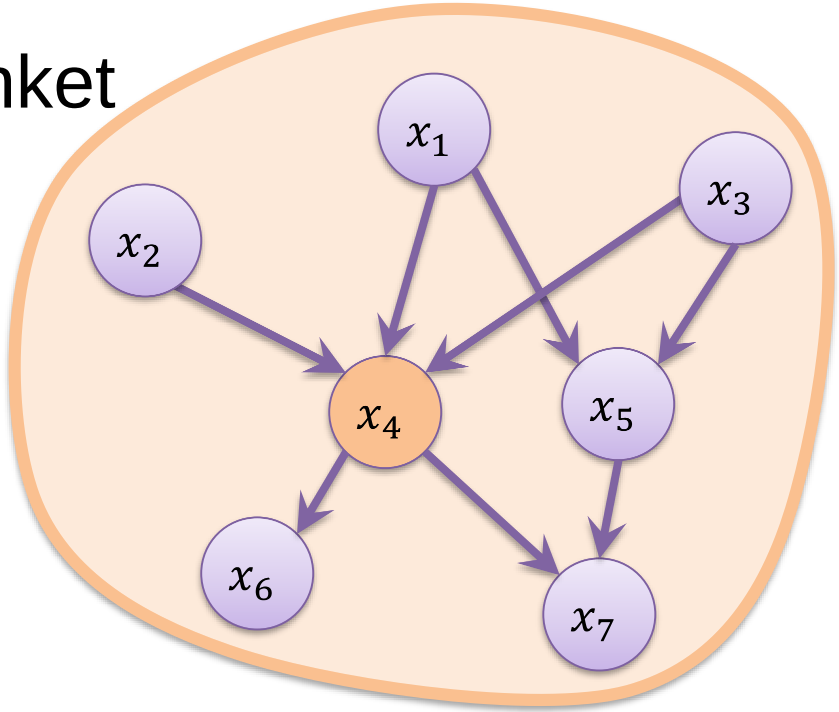
conditional independence

Markov blanket



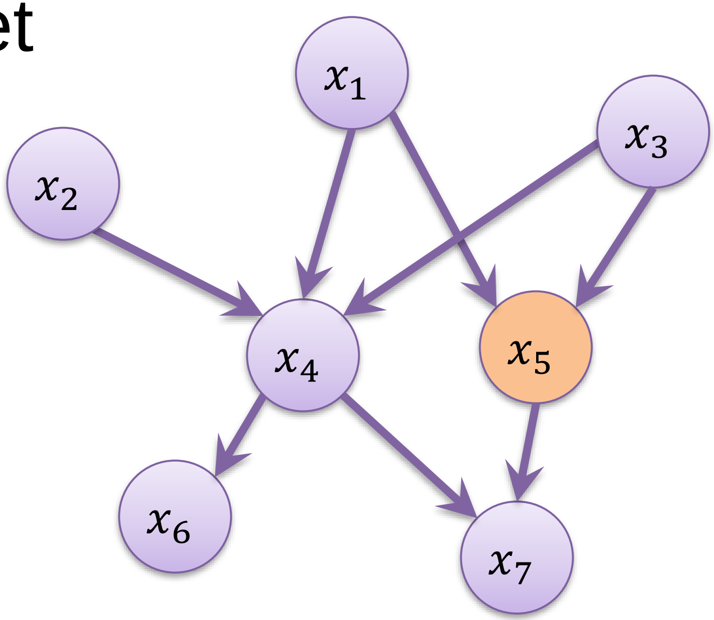
conditional independence

Markov blanket



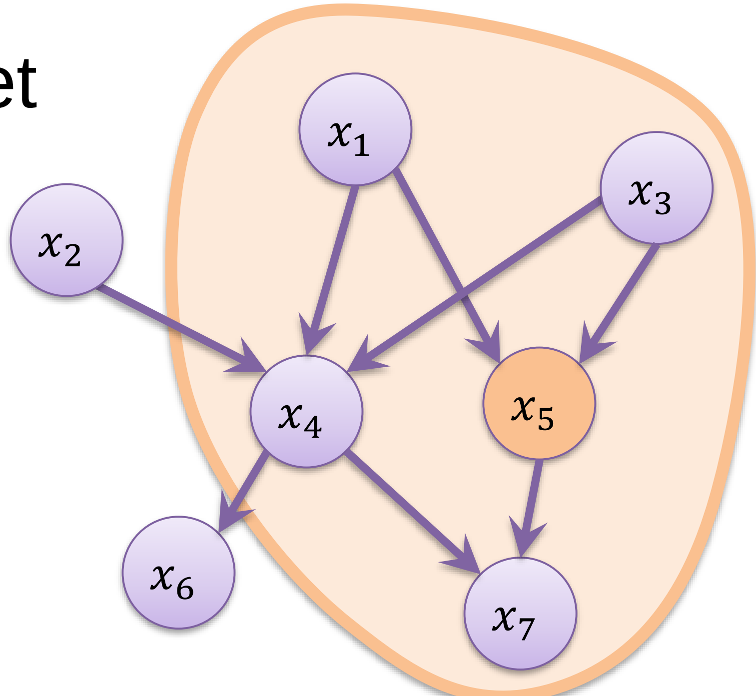
conditional independence

Markov blanket

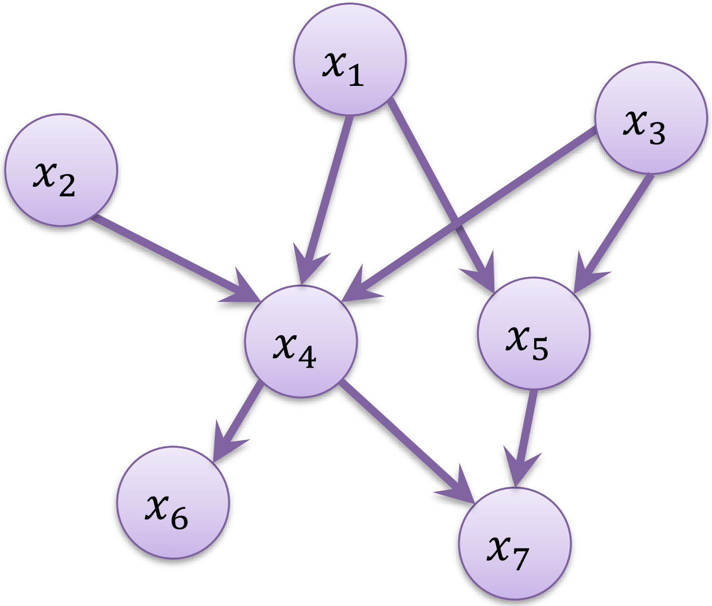


conditional independence

Markov blanket

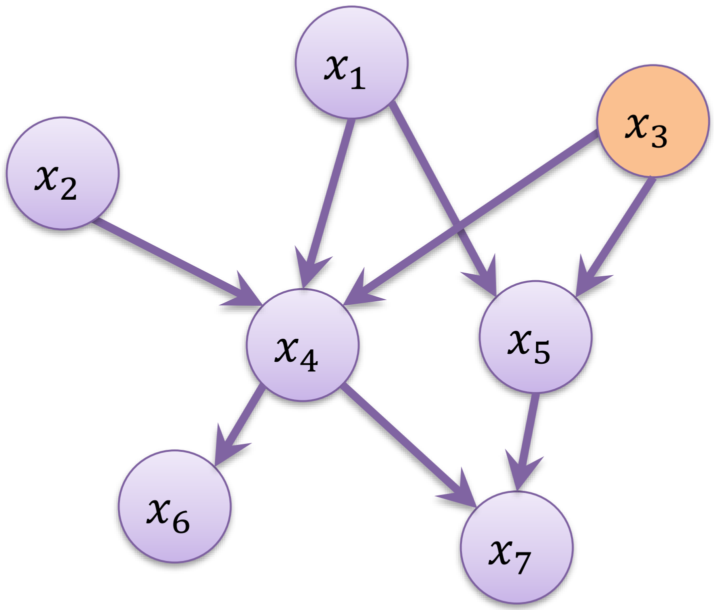


conditional independence



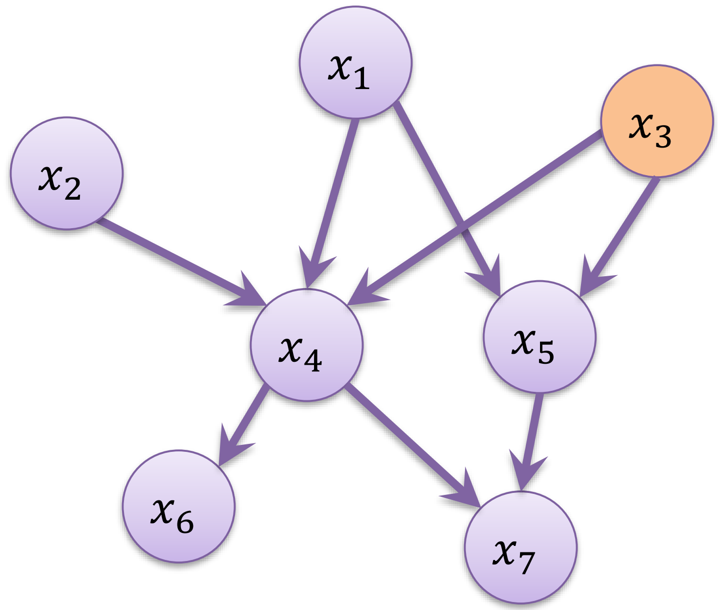
$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7) =$$
$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

conditional independence



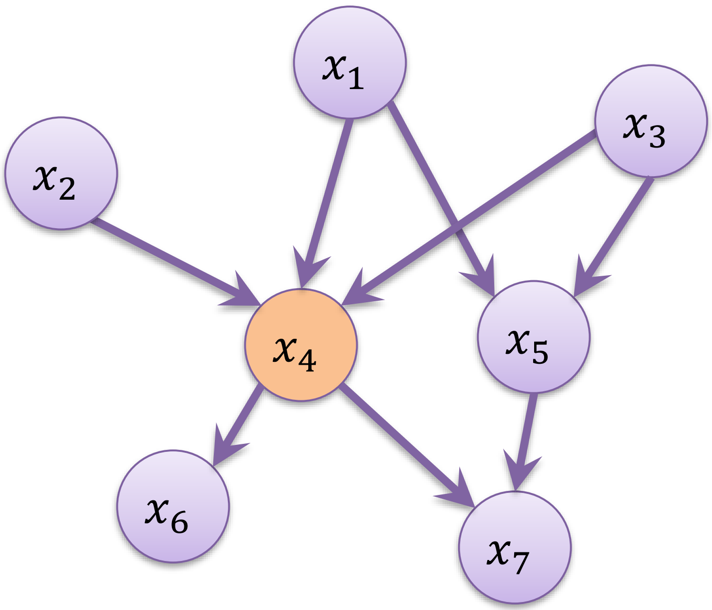
$$p(x_1, x_2, x_4, x_5, x_6, x_7 | x_3) =$$

conditional independence



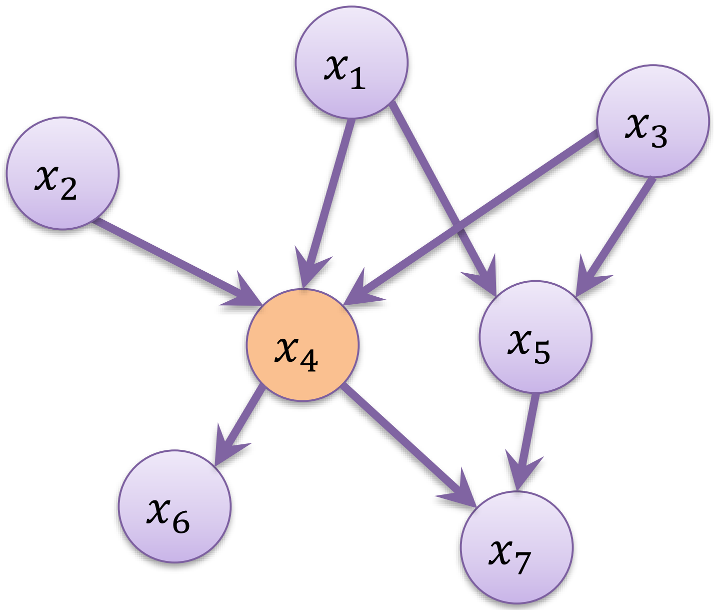
$$p(x_1, x_2, x_4, x_5, x_6, x_7 | x_3) =$$
$$p(x_1)p(x_2)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

conditional independence



$$p(x_1, x_2, x_3, x_5, x_6, x_7 | x_4) =$$

conditional independence



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

Resources

Bishop, C., Pattern Recognition and Machine Learning (2006)

MacKay, D., Information Theory, Inference, and Learning Algorithms (2005)

notes by Michael Jordan
notes by Zoubin Ghahramani