Introduction to Graphical Models
Exercices

Dr. Jean Marc Odobez (www.idiap.ch/∼odobez)

Outline

Outline

Party Animal

The problem

The party animal problem takes into account the following variables:

- P: been to party
- H: Got a headache
- D: Demotivated at work
- U: Underperform at work
- A: Boss angry

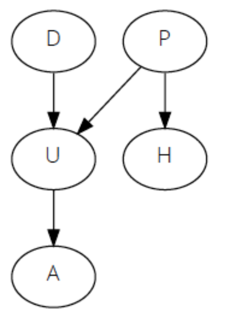
The distribution over these variables factorizes as:

$$p(P, H, D, U, A) = p(P)p(D)p(U|P, D)p(H|P)p(A|U)$$

Party Animal

a)Draw the graphical model corresponding to this problem.

$$p(P, H, D, U, A) = p(P)p(D)p(U|P, D)p(H|P)p(A|U)$$



Party Animal

b) Given that the boss is angry and that the worker has a headache, what is the probability that the worker has been to the party ?

We need to compute

$$p(P = T | A = T, H = T) = \frac{p(P = T, A = T, H = T)}{p(A = T, H = T)}$$

with

$$p(A = T, H = T) = p(P = T, A = T, H = T) + p(P = F, A = T, H = T)$$

We then just need to compute each required term.

Party Animal

b) Given that the boss is angry and that the worker has a headache, what is the probability that the worker has been to the party ?

$$p(P=T, A=T, H=T) = \sum_{u\in\{F,T\}, d\in\{F,T\}} p(P=T, A=T, H=T, U=u, D=d)$$

$p(P=T, A=T, H=T) =$

$$p(P=T)p(H=T|P=T) \sum_{u\in\{F,T\}} p(A=T|U=u) \sum_{d\in\{F,T\}} p(D=d)p(U=u|P=T, D=d)$$

We can show that:

$$\sum_{u\in\{F,T\}} p(A=T|U=u) \sum_{d\in\{F,T\}} p(D=d)p(U=u|P=T, D=d) = \mathbf{0.92282}$$

so that

$$p(P=T, A=T, H=T) = \mathbf{0.2} \times \mathbf{0.9} \times \mathbf{0.92282} = \mathbf{0.1661}$$

Party Animal

b) Given that the boss is angry and that the worker has a headache, what is the probability that the worker has been to the party ?

$$p(P=F, A=T, H=T) =$$

$$p(P=F)p(H=T|P=F)\sum_{u\in\{F,T\}} p(A=T|U=u)\sum_{d\in\{F,T\}} p(D=d)p(U=u|P=F, D=d)$$

We can show that:

$$\sum_{u\in\{F,T\}} p(A=T|U=u)\sum_{d\in\{F,T\}} p(D=d)p(U=u|P=F, D=d) = \mathbf{0.66227}$$

so that

$$p(P=F, A=T, H=T) = \mathbf{0.8} \times \mathbf{0.2} \times \mathbf{0.66227} = \mathbf{0.10596}$$

Party Animal

b) Given that the boss is angry and that the worker has a headache, what is the probability that the worker has been to the party ?

So the requested probability is:

$$p(P = T | A = T, H = T) = \frac{0.1661}{0.1661 + 0.10596} = 0.610$$

Outline

Asbestos

The problem

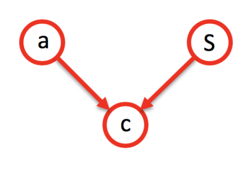
There is a synergistic relationship between Asbestos (A) exposure, Smoking (S) and Cancer (C). A model describing this relationship is given by

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

Is $A \perp\!\!\!\perp S$? Is $A \perp\!\!\!\perp S|C$?

Asbestos

Graph

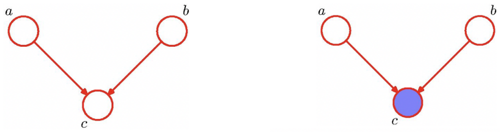


Is $A \perp\!\!\!\perp S$? Is $A \perp\!\!\!\perp S|C$?
cf course: 3 canonical graph

Asbestos

Graph

Is $A \perp\!\!\!\perp S$? Is $A \perp\!\!\!\perp S|C$?



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

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

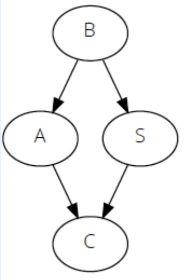
$$\underline{a \perp\!\!\!\perp b \mid \emptyset}$$

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

$$\underline{a \not\perp\!\!\!\perp b \mid c}$$

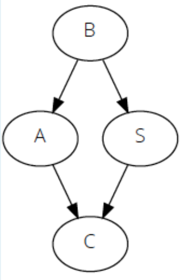
Asbestos

How could you adjust the model to account for the fact that people who work in the building industry have a higher likelihood to also be smokers and also a higher likelihood to be exposed to Asbestos ?



Asbestos

How could you adjust the model to account for the fact that people who work in the building industry have a higher likelihood to also be smokers and also a higher likelihood to be exposed to Asbestos ?



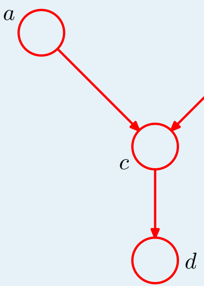
Adding the building node allows to have different statistics on whether people smoke or not (probability $P(S|B)$) or are exposed to asbestos (through the definition of $P(A|B)$).

Outline

Conditional independance

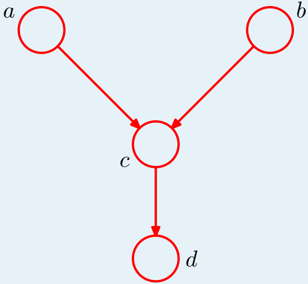
The problem

Consider the directed graphical model below, in which none of the variables is observed. Show that $a \perp\!\!\!\perp b$ (hint see how the demonstration was shown in the course for the canonical model 3).



Conditional independance

The problem : Show that $a \perp\!\!\!\perp b$



$$p(a,b) = \sum_{c,d} p(a,b,c,d) = \sum_{c,d} p(a)p(b)p(c|a,b)p(d|c)$$

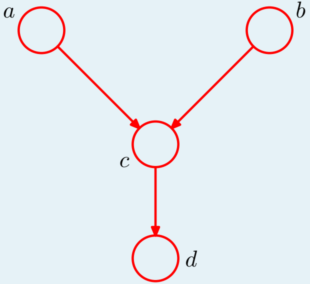
$$p(a,b) = p(a)p(b) \sum_{c} \left[p(c|a,b) \sum_{d} p(d|c) \right]$$

$$p(a,b) = \sum_{c,d} p(a,b,c,d) = p(a)p(b) \sum_{c} p(c|a,b) = p(a)p(b)$$

Conditional independance

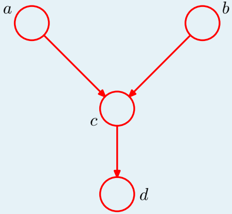
The problem

(Bonus. Not graded). Suppose now that we observe d. Show that in general $a \not\perp\!\!\!\perp b | d$.



Conditional independance

The problem



We want to show that in general $a \not\perp\!\!\!\perp b|d$. First note that if we would have had to show that:

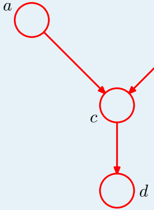
$$p \text{ is a distribution whose pdf factorizes according to the graph structure}$$
$$\Rightarrow a \perp\!\!\!\perp b|d$$

this would mean that we would have to show that

$$\text{for any distribution p, } (p \text{ respects the graph } \Rightarrow p(a, b|d) = p(a|d)p(b|d))\,.$$

Conditional independance

The problem



Thus, to show that the above does not hold, we have to show its negation which is defined as:

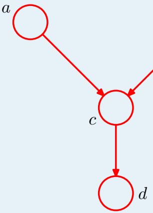
there exists at least one distribution such that

(p respects the graph AND the conditional independance does not hold) .

where the conditional independance does not hold means the observation of d makes the variable a and b dependent.

Conditional independance

The problem



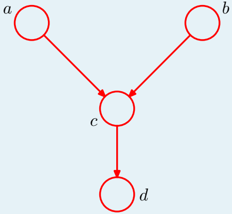
A class of such distribution is obtained by starting by using a distribution that follows the graph assumption (i.e. is defined by the local probability distributions), and for which we have (with c and d taking value in the same space),

$$p(d|c) = \delta_{c-d} \ .$$

where δ_x is equal to 1 when $x = 0$, and 0 otherwise. In other words, the probability of any (a, b, c, d) is zero whenever c and d are not the same. For such distributions, observing d is thus equivalent to observing c, which we know from the course makes the variables a and b dependent.

Conditional independance

More gneral case



Intuitively, what is happening in the more general case is that observing d provides information about the variable c (for instance, in the first exercice, look from the numerical values of $p(A|U)$ that the boss being angry bias the fact that the worker is underperforming towards true), or in other words, provides a soft evidence about the c value.

Outline

Chest Clinic Network

This network concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither). In this model, a visit to Asia is assumed to increase the probability of tuberculosis.

The involved variables are:

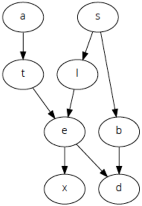
- x: positive X-ray
- d: Dyspnea (shortness of breath)
- e: either tuberculosis or lung cancer
- t: Tuberculosis
- l: Lung cancer
- b: Bronchitis
- a: visit to Asia
- s: Smoker

According to this model, the distribution factorizes as:

$$p(x, d, e, t, l, b, a, s) = p(a)p(s)p(t|a)p(l|s)p(b|s)p(d|b, e)p(e|t, l)p(x|e)$$

Chest Clinic Network

$$p(x, d, e, t, l, b, a, s) = p(a)p(s)p(t|a)p(l|s)p(b|s)p(d|b, e)p(e|t, l)p(x|e)$$



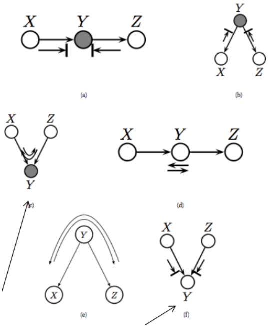
Next slides: figures shows in bold circles the nodes we are studying the conditional independance of - in blue: conditional variable (i.e. that are observed).

Note that in the questions, there could be an issue: when speaking about tuberculosis, we could wonder whether we should consider only the t node, or both the t and e nodes (since e is also characteristic of tuberculosis). Whichever you consider, the CI will be the same given their graph dependencies, so in practice, you may consider only the node t. The same applies for the lung cancer.

Chess Clinic Network

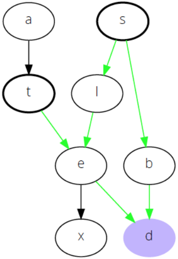
Reminder from the course: A path is blocked if two nodes on the path get independent, i.e. if it includes a node such that

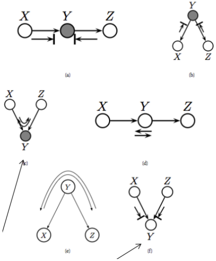
- the arrows on the path meet head-to-tail at the node AND the node is in C
- the arrows on the path meet tail-to-tail at the node AND the node is in C
- the arrows meet head-to-head at the node AND neither the node nor any of its descendants is in C



Chest Clinic Network

a) tuberculosis (t) $\perp\!\!\!\perp$ smoking (s) | shortness of breath (d) ?



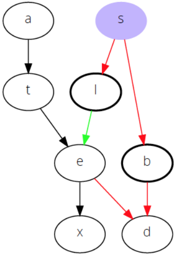


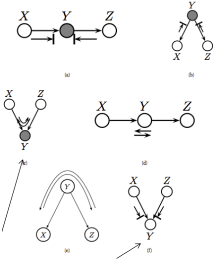
ANSWER : NOT TRUE.
There are two paths from node t to node s, none of them are blocking the path.

- For t-e-l-s, in e the arrow meet head-to-head and the observation d is in the descendant so does not bloc; in l, the arrow meets head-to-tail, not in d, so the path is not blocked.
- For t-e-d-b-s, e does not block, then head-to-head in d (not blocked), then in b, head-to-tail not in d. So this second path is not blocked.

Chest Clinic Network

b) lung cancer (l) $\perp\!\!\!\perp$ bronchitis (b) | smoking (s) ?



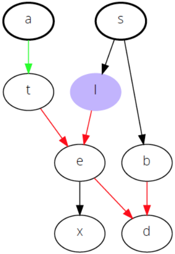


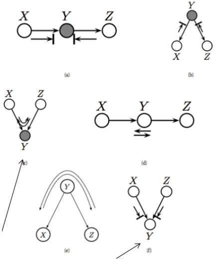
ANSWER: TRUE.
There are two paths from l to b.

- For l-s-b, we have a tail-to-tail in s, then it is blocked.
- For l-e-d-b, the path is blocked in d (head-to-head and node or descendant not in s).

Chest Clinic Network

c) visit to Asia (a) $\perp\!\!\!\perp$ smoking (s) | lung cancer (l) ?



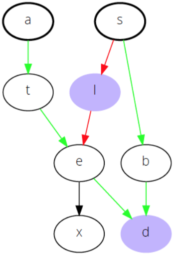


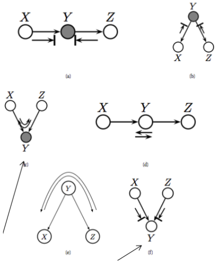
ANSWER: TRUE.
Two paths from a to s.

- a-t-e-l-s is blocked in e which is head-to-head with no descendent observed.
- Path a-t-e-d-b-s is blocked in d (head-to-head, and d not observed).

Chest Clinic Network

d) visit to Asia (a) $\perp\!\!\!\perp$ smoking (s) | lung cancer (l), shortness of breath (d)





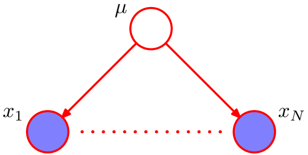
ANSWER: NOT TRUE.
There is a path which is not blocked (i.e. for which there is variable dependencies all along the chain). We have the same paths as in c).

- Path a-t-e-d-b-s is not blocked: in t, e, and b, we have tail-to-head nodes with middle node not observed, so the path is not blocked. In d, head-to-head node, with d observed, so the path is not blocked.

Outline

Gaussian mean

The problem



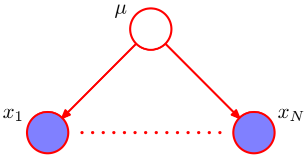
Consider the graphical model above, characterized by the following generative process:

$$\begin{aligned} \boldsymbol{\mu} &\sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \Sigma_0) && (1) \\ \mathrm{x}_i &\sim \mathcal{N}(\mathrm{x}|\boldsymbol{\mu}, \Sigma), \forall i && (2) \end{aligned}$$

and let us denote $\mathcal{X} = \{\mathrm{x}_i, i = 1 \ldots N\}$, and $\Theta = (\boldsymbol{\mu}_0, \Sigma_0, \Sigma)$.

Gaussian mean

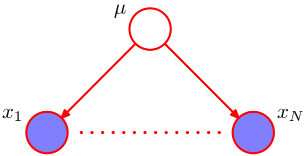
The problem

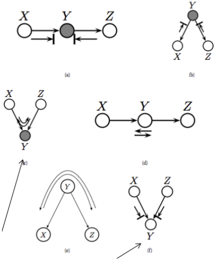


Using d-separation, state whether $x_i \perp\!\!\!\perp x_j \mid \boldsymbol{\mu}$ holds true or not.

Similarly, do we have $x_i \perp\!\!\!\perp x_j$ in general?

Gaussian mean

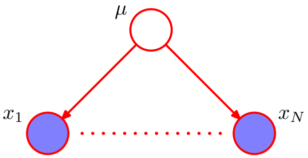




- When the mean is observed/given, we have a tail-to-tail node that is observed, so the x observations are independent, i.e. $x_i \perp\!\!\!\perp x_j \mid \boldsymbol{\mu}$.
- When the mean is not given, we have a tail-to-tail node that is not observed, and thus the observations are not independent.

Gaussian mean

Without computation and exploiting the course, state what is the type of the distribution of $p(\boldsymbol{\mu}|\mathcal{X})$?



cf. course: as all involved CPD are gaussians, then any distribution on the graph is gaussian, including $p(\boldsymbol{\mu}|\mathcal{X})$.

> *consider an arbitrary DAG over D variables, where each local CPD is expressed as a linear gaussian distribution. Then, the distribution over all components is a Gaussian.*
>
> *(slide 45)*

Gaussian mean

Suppose now that the x are one dimensionnal, so that $\boldsymbol{\mu}_0 = \mu_0$, $\Sigma_0 = \sigma_0$, $\Sigma = \sigma$. Compute analytically the parameters defining the distribution

$$p(\mu \mid \mathcal{X}) = p(\mu \mid x_1, \ldots, x_N)$$

The type of the distribution of $\mu|\mathcal{X}$ is:

$$\begin{aligned} p(\mu|\mathcal{X}) &= \frac{p(\mu, \mathcal{X})}{p(\mathcal{X})} && \text{Bayes' law} \\ &= \frac{p(\mu) \prod_n p(x_n|\mu)}{\prod_n p(x_n)} && \text{From the graph} \end{aligned}$$

$\prod_n p(x_n)$ is constant with respect to μ so

$$p(\mu|\mathcal{X}) \propto p(\mu) \prod_n p(x_n|\mu)$$

so $p(\mu|\mathcal{X})$ is a gaussian distribution as a product of gaussian distributions.

Gaussian mean

We have

$$p(\mu)\prod_n p(x_n|\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{1}{2\sigma_0}(\mu-\mu_0)^2\right]\prod_n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma}(x_i-\mu)^2\right]$$
$$= \frac{1}{\sqrt{2\pi\sigma_0^2}}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^N \exp\left[-\frac{1}{2\sigma_0}(\mu-\mu_0)^2 - \sum_n \frac{1}{2\sigma}(x_i-\mu)^2\right]$$

We can find the mean of this distribution f by solving

$$\frac{\partial f}{\partial \mu} = 0$$

Since $(ke^{u(t)}))' = ku'(t)e^{u(t)}$, solving $(ke^{u(t)}))' = 0$ is equivalent to solving $u'(t) = 0$. So here, we solve

$$\frac{\partial}{\partial \mu}\left[-\frac{1}{2\sigma_0}(\mu-\mu_0)^2 - \sum_n \frac{1}{2\sigma}(x_i-\mu)^2\right] = 0$$

Gaussian mean

We have

$$-\frac{1}{\sigma_0}(\mu - \mu_0) + \frac{1}{\sigma}\sum_n (x_i - \mu) = 0$$

which lead to

$$\mu = \frac{\sigma_0 \sum_n x_i + \sigma \mu_0}{\sigma + N\sigma_0}$$

Thank you for your attention!

Dr. Jean Marc Odobez (www.idiap.ch/∼odobez)
Idiap Research Institute, Martigny, Switzerland