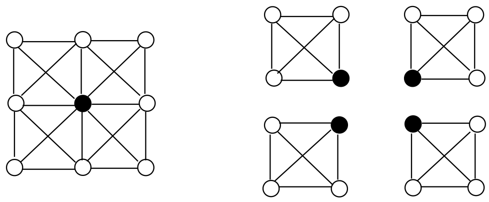
In what follows [Dav] refers to Davison's book, and [Jen] refers to Jensen's 1996 book on Bayesian networks.

1. Do Exercises 6.2.2, 6.2.4 in [Dav] (p. 254).

Solution: (i) Ex. 6.2.2: In the second-order neighboring system each site has 8 neighbors: 2 adjacent horizontal sites, 2 adjacent vertical sites, and 4 adjacent diagonal sites. In the figure below, the larger graph on the left is the subgraph induced by a site (black) and its neighbors (white), and there are four cliques containing the site, illustrated by the four smaller graphs on the right.



(ii) Ex. 6.2.4: The local characteristics are identical to

$$P(Y_1 = 0 \,|\, Y_2 = 0) = P(Y_1 = 1 \,|\, Y_2 = 1) = P(Y_2 = 1 \,|\, Y_1 = 0) = P(Y_2 = 0 \,|\, Y_1 = 1) = 0,$$

which implies that no pairs of values of Y_1 and Y_2 can occur jointly. So there does not exist a joint distribution P satisfying these local characteristics, and the positivity condition is of course violated. □

2. Do Exercise 6.2.3 in [Dav] (p. 254). Also answer the following question for the undirected graph associated with a first-order Markov chain: Suppose that $P(X_1, \dots X_n)$ is a homogeneous MRF with respect to that graph, and that P satisfies the positivity condition. Is $(X_1, \dots, X_n)$ also a homogeneous first-order Markov chain?

Solution: (i) Ex. 6.2.3: For a second-order Markov chain $Y = (Y_1, \dots, Y_n)$ of length n, the cliques are

$$\{i-2\,,\, i-1\,,\, i\}, \quad 3 \le i \le n.$$

So, under the positivity condition, $p(y) \propto \exp\{-\psi(y)\}$, and in the most general form, $\psi(y)$ can be expressed as

$$\psi(y) = a + \sum_{i=1}^{n} b_i(x_i) + \sum_{i=2}^{n} c_i(x_{i-1}, x_i) + \sum_{i=3}^{n} d_i(x_{i-2}, x_i) + \sum_{i=3}^{n} e_i(x_{i-2}, x_{i-1}, x_i)$$

for some constant a and functions b_i, c_i, d_i, e_i.

(ii) No, $(X_1, \dots, X_n)$ corresponding to a homogeneous MRF is an inhomogeneous Markov chain, as can be verified from the expressions of $P(X_n \,|\, X_{n-1}), P(X_{n-1} \,|\, X_{n-2}), \dots$ directly. Let $n \ge 3$ and $X = (X_1, \dots, X_n)$. Suppose

$$p(x) \propto \exp\Big\{ -\sum_{i=1}^{n} b(x_i) - \sum_{i=2}^{n} c(x_{i-1}, x_i) \Big\}.$$

For notational simplicity, define a function $\phi(x_{i-1}, x_i) = -b(x_i) - c(x_{i-1}, x_i)$. Then, for fixed x_{n-1},

$$p(x_n \,|\, x_{n-1}) \propto \exp\big\{\phi(x_{n-1}, x_n)\big\},$$

while for fixed x_{n-2}, we have

$$\begin{aligned} p(x_{n-1} \,|\, x_{n-2}) = \sum_{x_n} p(x_{n-1}, x_n \,|\, x_{n-2}) &\propto \sum_{x_n} \exp\big\{\phi(x_{n-2}, x_{n-1}) + \phi(x_{n-1}, x_n)\big\} \\ &= \exp\big\{\phi(x_{n-2}, x_{n-1})\big\} \cdot \sum_{x_n} \exp\big\{\phi(x_{n-1}, x_n)\big\}. \end{aligned}$$

Comparing the two expressions above, we see that the transition probabilities depend on the time step. (This relates to the discussion on CRF vs. HMM in one of the lecture.) □

3. Do Exercise 3.1 in [Jen] (p. 64).

Solution: (i) For the model that has Ho as the mediating variable, we have

$$\begin{aligned}P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n) &= \sum_{x\in\{y,n\}} P(\texttt{BT}=n \mid \texttt{Ho}=x)P(\texttt{UT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=n)P(\texttt{Pr}=n)\\ &= 0.3\cdot 0.2\cdot 0.01\cdot 0.13 + 0.9\cdot 0.9\cdot 0.99\cdot 0.13 \approx 0.1043,\\ P(\texttt{Pr}=y,\texttt{BT}=n,\texttt{UT}=n) &= \sum_{x\in\{y,n\}} P(\texttt{BT}=n \mid \texttt{Ho}=x)P(\texttt{UT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=y)P(\texttt{Pr}=y)\\ &= 0.3\cdot 0.2\cdot 0.9\cdot 0.87 + 0.9\cdot 0.9\cdot 0.1\cdot 0.87 \approx 0.1175.\end{aligned}$$

So

$$\begin{aligned}P(\texttt{Pr}=n \mid \texttt{BT}=n,\texttt{UT}=n) &= \frac{P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n)}{P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n)+P(\texttt{Pr}=y,\texttt{BT}=n,\texttt{UT}=n)}\\ &\approx \frac{0.1043}{0.1043+0.1175} \approx 0.47.\end{aligned}$$

(ii) Under the same model as above, we have

$$\begin{aligned}P(\texttt{BT}=n \mid \texttt{Pr}=n) &= \sum_{x\in\{y,n\}} P(\texttt{BT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=n) = 0.894,\\ P(\texttt{BT}=n \mid \texttt{Pr}=y) &= \sum_{x\in\{y,n\}} P(\texttt{BT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=y) = 0.36,\end{aligned}$$

and

$$\begin{aligned}P(\texttt{UT}=n \mid \texttt{Pr}=n) &= \sum_{x\in\{y,n\}} P(\texttt{UT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=n) = 0.893,\\ P(\texttt{UT}=n \mid \texttt{Pr}=y) &= \sum_{x\in\{y,n\}} P(\texttt{UT}=n \mid \texttt{Ho}=x)P(\texttt{Ho}=x \mid \texttt{Pr}=y) = 0.27.\end{aligned}$$

Using these as parameters for the naive Bayes model, we have

$$\begin{aligned}P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n) &= P(\texttt{BT}=n \mid \texttt{Pr}=n)P(\texttt{UT}=n \mid \texttt{Pr}=n)P(\texttt{Pr}=n) = 0.894\cdot 0.893\cdot 0.13,\\ P(\texttt{Pr}=y,\texttt{BT}=n,\texttt{UT}=n) &= P(\texttt{BT}=n \mid \texttt{Pr}=y)P(\texttt{UT}=n \mid \texttt{Pr}=y)P(\texttt{Pr}=y) = 0.36\cdot 0.27\cdot 0.87,\end{aligned}$$

so

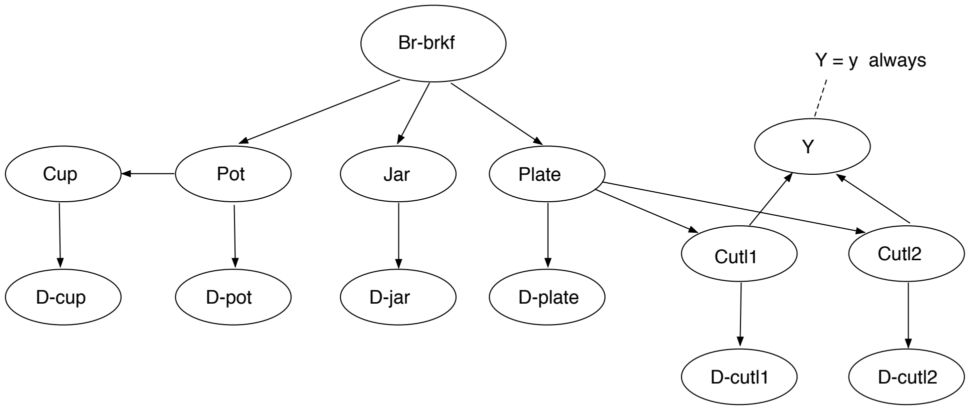
$$\begin{aligned}P(\texttt{Pr}=n \mid \texttt{BT}=n,\texttt{UT}=n) &\frac{P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n)}{P(\texttt{Pr}=n,\texttt{BT}=n,\texttt{UT}=n)+P(\texttt{Pr}=y,\texttt{BT}=n,\texttt{UT}=n)}\\ &= \frac{0.894\cdot 0.893\cdot 0.13}{0.894\cdot 0.893\cdot 0.13 + 0.36\cdot 0.27\cdot 0.87} \approx 0.55\end{aligned}$$

under the naive Bayes model. □

4. Do Exercise 3.15 in [Jen] (p. 66-67).

Solution: I describe first a model that seems reasonable to me, and then another model from a participant of the exercise group, which seems more intuitive than mine. You can have a different answer provided that you can justify it. Besides practicing model building, a particular point to pay attention to in this exercise is to take care of the dependences/constraints between the variables.

My model is shown in the figure below. The hypothesis variable is British-breakfast (Br-brkf), which takes two values, yes (y) and no (n). The information variables are the variables at the bottom level of the DAG: D-cup, D-pot, D-jar, etc., and they correspond to the detected type of the cup, pot, jar, etc., respectively. The rest are mediating variables: Cup represents the true type of the cup, Pot that of the pot, etc., and the variable Y is introduced to constrain the possible pairs of cutlery that can be present on the table. The values that the variables can take and the conditional probabilities for the model components are specified in the tables below with explanations.



Br-brkf, Pot, Cup:

Br-brkf	y	n
	0.5	0.5

Br-brkf \ Pot	tea	coffee
y	0.99	0.01
n	0	1

Pot \ Cup	tea	coffee
tea	1	0
coffee	0	1

I let Cup be the child of Pot to encode the constraint that a tea (coffee) pot should match a tea (coffee) cup. (Perhaps $P(\texttt{Pot} = \text{coffee} \mid \texttt{Br-brkf} = y)$ could be made smaller to match the description of the exercise better; but I am not worried about this.)

Jar, Plate:

Br-brkf \ Jar	orange	red
y	0.99	0.01
n	0	1

Br-brkf \ Plate	big	small
y	0.7	0.3
n	0.5	0.5

In the above, I assumed that big plates are usually used in British breakfast, while both types of plates are equally likely for continental breakfast.

Plate, Cutlery-1 (Cutl1), Cutlery-2 (Cutl2):

From the description of the exercise, we know that the two pieces of cutlery present on the table can only be knife-fork or knife-spoon. I assume that which combination appears depends on the type of the plate used, in particular,

$$\text{Prob(knife and fork are present} \mid \texttt{Plate} = \text{big}) = 0.8, \tag{1}$$

$$\text{Prob(knife and spoon are present} \mid \texttt{Plate} = \text{small}) = 0.8. \tag{2}$$

In the model, Cutl1 and Cutl2 are independent given Plate. I want to put an undirected edge between Cutl1 and Cutl2, encoding the constraint on the knife-fork and knife-spoon combinations, as well as setting the conditional probabilities to match the desired probabilities given in Eqs. (1)-(2). (Another way is to put a directed edge; see the other model at the end.) For this purpose, I introduce the mediating variable Y which will be set at $Y = y$ always. The constraint is encoded in the conditional probabilities for $Y = y$, as shown in the right table below.

Now I need to specify the conditional probabilities for Cutl1 and Cutl2 given Plate. Suppose that these are the same for Cutl1 and Cutl2, and for $x = \text{knift}, \text{spoon}, \text{fork}$, $P(\texttt{Cutl1} = x \mid \texttt{Plate} = \text{big})$ is $\alpha_1, \alpha_2, \alpha_3$, respectively, while $P(\texttt{Cutl1} = x \mid \texttt{Plate} = \text{small})$ is $\alpha'_1, \alpha'_2, \alpha'_3$, respectively. Then

$$\frac{P(\text{knife and fork are present} \mid \texttt{Plate} = \text{big}, Y = y)}{P(\text{knife and spoon are present} \mid \texttt{Plate} = \text{big}, Y = y)} = \frac{2\alpha_1\alpha_3}{2\alpha_1\alpha_2} = \alpha_3/\alpha_2.$$

So, in order to match Eq. (1), I only need to have $\alpha_3/\alpha_2 = 0.8/0.2 = 4$. Similarly, for the case $\texttt{Plate} = \text{small}$, in order to match Eq. (2), $\alpha'_2/\alpha'_3 = 0.8/0.2 = 4$. The absolute values of α_1, α'_1 do not matter, but it must be $\alpha_1 = \alpha'_1$, otherwise from $Y = y$ there would be a bias in the type of the plate used. This explains the numbers in the left table below.

Plate \ Cutl1/Cutl2	knife	spoon	fork
big	0.5	0.1	0.4
small	0.5	0.4	0.1

$P(Y = y\,|\,\texttt{Cult1}, \texttt{Cult2})$:

Cutl1 \ Cutl2	knife	spoon	fork
knife	0	1	1
spoon	1	0	0
fork	1	0	0

For the detected types of the objects:

Pot \ D-pot	tea	coffee
tea	0.6	0.4
coffee	0.4	0.6

Cup \ D-cup	tea	coffee
tea	0.7	0.3
coffee	0.2	0.8

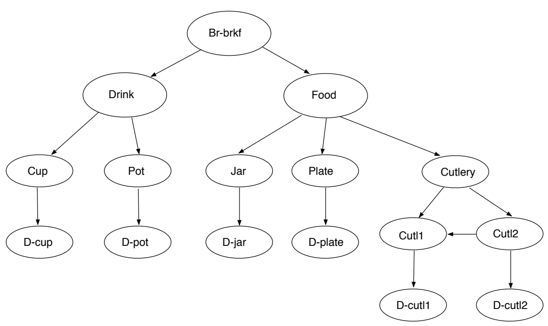
Jar \ D-jar	orange	red
orange	0.95	0.05
red	0.05	0.95

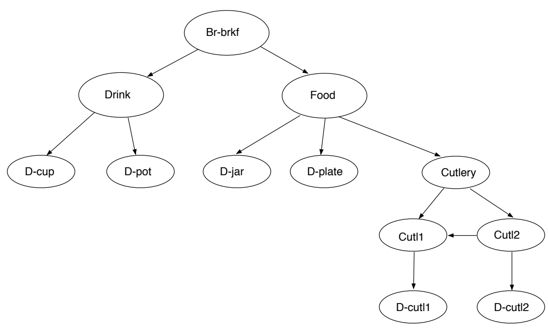
Plate \ D-plate	big	small
big	0.9	0.1
small	0.1	0.9

Cutl1 \ D-cutl1	knife	spoon	fork
knife	0.85	0.05	0.1
spoon	0	0.75	0.25
fork	0.1	0.2	0.7

These probabilities are as given by the description of the exercise.

Another model:





We can start with the model on the left, and then marginalize out the mediating variables Cup, Pot, etc., to obtain the simplified model on the right with the conditional probabilities calculated accordingly. (Cutl1 and Cutl2 may also be marginalized out if one wishes so.) □

5. Think about a problem – of any kind, not necessarily research-related – that can be modeled using either MRF or Bayesian networks. Give an description of the problem and specify the model: random variables, the graph structure, and how you would assign probabilities, as well as any decisions you made in choosing a particular model. Describe also what inference tasks you would like to perform using the model.