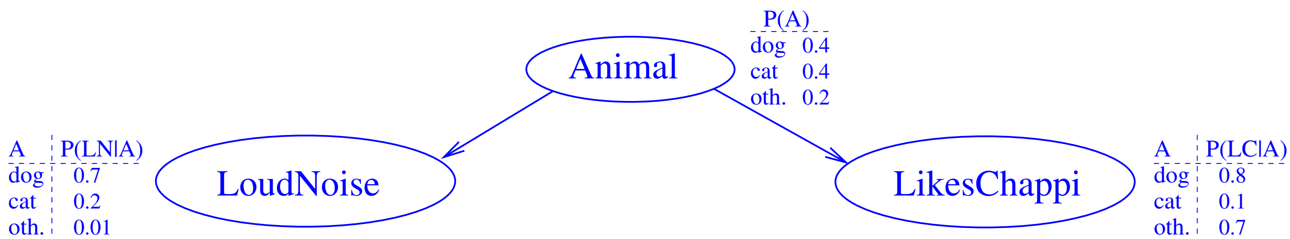
Exercise 1 :

Consider the following Bayesian network BN:



Use inference by enumeration to compute the following probabilities:

(a) $P(dog \mid loudnoise, likeschappi)$.

(b) $P(loudnoise \mid \neg likeschappi)$.

Include intermediate steps at a level of granularity as in the lecture slides examples. In particular, for each of (a) and (b), state what the query variable, evidence, and hidden variables are; and write down *which* probabilities provided in BN can be combined *how* to obtain the demanded probability P.

Solution:

As the variable ordering consistent with BN, we choose $X_1 = Animal$, $X_2 = LoudNoise$, $X_3 = LikesChappi$.

(a) The query variable X here is $Animal$. The evidence $\mathbf{e}$ is $loudnoise, likeschappi$. There are no hidden variables, $\mathbf{Y} = \emptyset$. Using Normalization+Marginalization $\mathbf{P}(X \mid \mathbf{e}) = \alpha \sum_{\mathbf{y} \in \mathbf{Y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$ we get $\mathbf{P}(Animal \mid loudnoise, likeschappi) = \alpha \mathbf{P}(Animal, loudnoise, likeschappi)$. By the Chain rule and exploiting conditional independence, we get $\alpha \mathbf{P}(Animal, loudnoise, likeschappi) = \alpha \mathbf{P}(likeschappi \mid Animal) * \mathbf{P}(loudnoise \mid Animal) * \mathbf{P}(Animal) = \alpha \langle 0.8 * 0.4 * 0.7, 0.1 * 0.2 * 0.4, 0.7 * 0.01 * 0.2 \rangle \approx \langle 0.96, 0.03, 0.01 \rangle$. Thus $P(dog \mid loudnoise, likeschappi) \approx 0.96$. ("If your animal likes Chappi and makes loud noise, chances are good it's a dog.")

(b) The query variable X here is $LoudNoise$. The evidence $\mathbf{e}$ is $\neg likeschappi$. The hidden variables $\mathbf{Y}$ are $\{Animal\}$. Using Normalization+Marginalization $\mathbf{P}(X \mid \mathbf{e}) = \alpha \sum_{\mathbf{y} \in \mathbf{Y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$ we get $\mathbf{P}(LoudNoise \mid \neg likeschappi) = \alpha \sum_{\mathbf{y} \in \{Animal\}} \mathbf{P}(\mathbf{y}, LoudNoise, \neg likeschappi)$. By the Chain rule and exploiting conditional independence, we get $\alpha \sum_{\mathbf{y} \in \{Animal\}} \mathbf{P}(\mathbf{y}, LoudNoise, \neg likeschappi) = \alpha \sum_{\mathbf{y} \in \{Animal\}} P(\neg likeschappi \mid \mathbf{y}) * \mathbf{P}(LoudNoise \mid \mathbf{y}) * P(\mathbf{y})$. That is, for each truth value v of $LoudNoise$, we need to sum over each possible animal $\mathbf{y}$ the product of the likelihood of $\mathbf{y}$ not liking Chappi multiplied with the likelihood of $\mathbf{y}$ having value v for $LoudNoise$ multiplied with the likelihood of $\mathbf{y}$. This gives us $\mathbf{P}(LoudNoise \mid \neg likeschappi) = \alpha \langle 0.2*0.7*0.4 + 0.9*0.2*0.4 + 0.3*0.01*0.2, 0.2*0.3*0.4 + 0.9*0.8*0.4 + 0.3*0.99*0.2 \rangle = \alpha \langle 0.1286, 0.3714 \rangle = \langle 0.2572, 0.7428 \rangle$. Thus $P(loudnoise \mid \neg likeschappi) = 0.2572$. ("If your animal does not like Chappi, chances are good it is quiet.")

Exercise 2 :

Consider the running example of the lecture with the following probability tables:

	Burglary	
	t	f
	.1	.9

	Earthquake	
	t	f
	.1	.9

		Alarm	
Burglary	Earthquake	t	f
t	t	.9	.1
t	f	.8	.2
f	t	.5	.5
f	f	.1	.9

	JohnCalls	
Alarm	t	f
t	.8	.2
f	.1	.9

	MaryCalls	
Alarm	t	f
t	.7	.3
f	.1	.9

Use Variable Elimination to determine the conditional probability $P(MC \mid B = t)$. For each step indicate the operations that are performed over the previous factors and what new factor is computed.

Solution:

The abstract computation on slide 6.16 are given as:

$$\sum_{a\in\{t,f\}}\sum_{eq\in\{t,f\}}\sum_{jc\in\{t,f\}} P(B=t)P(EQ=eq)P(A=a \mid B=t, EQ=eq)P(JC=jc \mid A=a)P(MC \mid A=a) =$$
$$\sum_{a\in\{t,f\}}\sum_{eq\in\{t,f\}} P(B=t)P(EQ=eq)P(A=a \mid B=t, EQ=eq)P(MC \mid A=a)F_1(a) =$$
$$\sum_{a\in\{t,f\}} P(B=t)P(MC \mid A=a)F_1(a)F_2(a) =$$
$$P(B=t)F_3(MC)$$

With the concrete tables we have specified for the conditional distributions, we can now compute: to eliminate the variable JohnCalls, we multiply all tables that contain the variable JohnCalls, and sum out the JohnCalls variable. The result is the table, or factor, F_1. Since there is only one table containing JohnCalls, the result is very simple:

	$F_1(Alarm)$	
	t	f
	1	1

Next we eliminate *Earthquake*. There are two tables containing Earthquake: the table of the Earthquake node, and the conditional distribution of Alarm. The latter table is restricted to the cases that are consistent with the observed evidence $B = t$, which gives a table only containing Earthquake and Alarm:

	Alarm	
Earthquake	*t*	*f*
t	.9	.1
f	.8	.2

Multiplying this with the Earthquake table gives

	Alarm	
Earthquake	*t*	*f*
t	.09	.01
f	.72	.18

Summing out the Earthquake variable then gives the factor F_2:

	$F_2(Alarm)$	
	t	*f*
	0.81	0.19

Finally, we eliminate *Alarm*. Multiplying F_1, F_2, and the table for MaryCalls gives

	MaryCalls	
Alarm	*t*	*f*
t	.7· 0.81=0.567	.3 · 0.81=0.243
f	.1· 0.19 = 0.019	.9 · 0.19= 0.171

Summing out Alarm gives

	$F_3(MaryCalls)$	
	t	*f*
	0.586	0.414

This multiplied with $0.1 = P(B = t)$ gives

	$P(B = t)F_3(MaryCalls)$	
	t	*f*
	0.0586	0.0414

which is now the table containing the function $P(B = t, MC)$. To obtain the conditional distribution $P(MC \mid B = t)$ this has only to be normalized by dividing with 0.0586+0.0414, which gives

	$P(MaryCalls \mid B = t)$	
	t	*f*
	0.586	0.414

Note that by inspecting the semantics of the potentials calculated above, we could have stopped after summing out Alarm, since the resulting potential is the conditional distribution $P(MC \mid B = t)$.

Exercise 3 :

* Complete Exercise 8.10 in PM.

Solution: To calculate $P(E)$ you can start by removing D and F, since they are both barren, i.e., the result of marginalizing out these two variables will simply be unity factors. Thus, we end up with $P(A)$, $P(B)$, $P(C|A,B)$, and $P(E|C)$.

Start by eliminating A:

This will create a factor $F_1(C,B) = \sum_A P(A)P(C|A,B)$, with the intermediate factor $P(A)P(C|A,B)$

		C	
A	B	t	f
t	t	0.09	0.81
t	f	0.72	0.18
f	t	0.07	0.03
f	f	0.04	0.06

and the final factor

	C	
B	t	f
t	0.16	0.84
f	0.76	0.24

Eliminate B:

This will create a factor $F_2(C) = \sum_B P(B)F_1(C|B) = (0.64, 0.36)$.

Eliminate C:

This will create a factor $F_3(E) = \sum_C P(E|C)F_2(C) = (0.52, 0.48)$.

Exercise 4 :

Consider the network defined by the two binary variables A and B, where A is the parent of B. Assume that the conditional probability tables are given as $P(A) = (0.1, 0.9)$ and

	A	
	a_1	a_2
b_1	0.05	0.2
b_2	0.95	0.8

i) Assume that you want to estimate $P(b_1)$ using sampling. How many samples would be required if you only accept an error larger than 0.15 in 10% of the cases?

ii) Generate two random samples, using the following list of random numbers (any time that you use a random number pick one from the list): $0.5, 0.3, 0.01, 0.8$.

iii) Suppose that after sampling 10 states, you got $(A = a_2, B = b_2)$ 3 times, and $(A = a_1, B = b_2)$,6 times, and $(A = a_1, B = b_1)$ once. Estimate $P(B = b_1)$

iv) (Optional) Implement the network above in Hugin and use Hugin to sample the number of cases that you calculated in step (i); use the function 'Simulate cases' under 'File'. Use the sampled cases to estimate $P(b_1)$ and compare the result with Hugin. Feel free to use a spreadsheet for the counting.

v) Assume that you want to use rejection sampling to estimate $P(A|B = b_1)$. How many samples do you expect you would have to generate in order to end up (after rejection) with a sample set of 1000 cases for estimating the probability.

Solution:

i) Hoeffding's inequality gives us

$$P(|s - p| > 0.15) \leq 2e^{-2n0.15^2} < 0.1,$$

which we can rewrite to find that $n > 66.57$.

ii) As A is parent of B, we always sample A first.

First state: $A = a_2, B = b_2$: $0.5 > 0.1$, so $A = a_2$, then $0.3 > 0.2$ so $B = b_2$. Second state: $A = a_1, B = b_2$: $0.01 \leq 0.1$, so $A = a_1$, then $0.8 > 0.0.05$ so $B = b_2$.

iii) $P(B = b_1) = 0.1$

iv) (Must be done in Hugin)

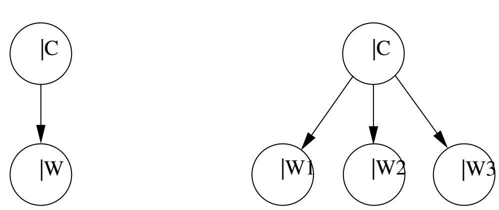
v) The probability of $B = b_1$ is $P(B = b_1) = 0.185$, hence only 18.5% of all the generated samples would not be rejected. We therefore need to generate approximately 5400 cases to end up with a sample set of 1000 cases.

Exercise 5 :

Complete Exercise 8.6(a-b) in PM.

Solution:

We have two variables in the network: *Company* (C) with states *green* (g) and *blue* (b) and *Witness* (W) with states *green* (g) and *blue* (b). The structure of the network incorporating one witness is illustrated in Figure 1(a).



(a) (b)

Figure 1: The witness model: Model (a) includes one witness and model (b) includes three witnesses.

The conditional probability tables for the model are given by:

$P(C=g)$	$P(C=b)$
0.85	0.15

	$C=g$	$C=b$
$W=g$	0.8	0.2
$W=b$	0.2	0.8

$P(W \mid C)$

For calculating $P(C=b \mid W=b)$ we use Bayes rule:

$$\begin{aligned} P(C=b \mid W=b) &= \frac{P(W=b \mid C=b)P(C=b)}{P(W=b)} \\ &= \frac{P(W=b \mid C=b)P(C=b)}{P(C=b, W=b) + P(C=g, W=b)} \\ &= \frac{0.8 \cdot 0.15}{0.8 \cdot 0.15 + 0.2 \cdot 0.85} \\ &= \frac{0.12}{0.12 + 0.17} \approx 0.41 \end{aligned}$$

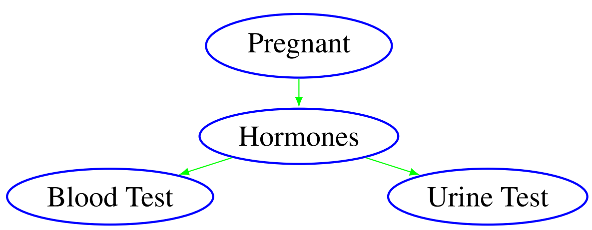
With three independent witnesses we have the network in Figure 1(b), where we assign the same conditional probability distribution to each of the three witnesses. Thus, we have a naive Bayes model with three information variables.

We can perform probability updating in this model as shown on the slides in the previous lecture. E.g.

$$\begin{aligned} P(C=b \mid W_1=W_2=W_3=b) &= \frac{P(W_1=b \mid C=b)P(W_2=b \mid C=b)P(W_3=b \mid C=b)P(C=b)}{P(W_1=b, W_2=b, W_3=b)} \\ &= \frac{0.8 \cdot 0.8 \cdot 0.8 \cdot 0.15}{0.8 \cdot 0.8 \cdot 0.8 \cdot 0.15 + 0.2 \cdot 0.2 \cdot 0.2 \cdot 0.85} = 0.92 \end{aligned}$$

Exercise 6 :

Consider the insemination example from Section 3.1.13 in BNDG:



Let the probabilities be as in Table 1 ($Ho = y$ means that hormonal changes have taken place) $P(Pr) = (0.87, 0.13)$.

	$Pr = y$	$Pr = n$
$Ho = y$	0.9	0.01
$Ho = n$	0.1	0.99

	$Ho = y$	$Ho = n$
$BT = y$	0.7	0.1
$BT = n$	0.3	0.9

	$Ho = y$	$Ho = n$
$UT = y$	0.8	0.1
$UT = n$	0.2	0.9

Table 1: Tables for Exercise 6.

i) What is $P(Pr \mid BT = n, UT = n)$?

ii) Construct a naive Bayes model. Determine the conditional probabilities for the model by making inference queries in the model above using Hugin. What is $P(Pr \mid BT = n, UT = n)$ in this model and how does it compare to the result you got above? Try to (qualitatively) account for any differences.

iii) (Optional) Verify your solution modelling both BNs in Hugin.

Solution: *Part 1:*

$$P(Pr|Bt = n, Ut = n) = \frac{P(Pr, Bt = n, Ut = n)}{P(Bt = n, Ut = n)}$$

Next, we should calculate $P(Pr, Bt = n, Ut = n)$ and $P(Bt = n, Ut = n)$:

$$P(Pr, Bt = n, Ut = n) = \sum_{Ho} P(Pr, Bt = n, Ut = n, Ho)$$
$$P(Bt = n, Ut = n) = \sum_{Pr} P(Pr, Bt = n, Ut = n)$$

Thus, we only need to calculate $P(Pr, Bt = n, Ut = n, Ho)$ and this can be done using the chain rule:

$$P(Pr, Bt = n, Ut = n, Ho) = P(Ut = n|Ho)P(Bt = n|Ho)P(Ho|Pr)P(Pr)$$

The final value is $P(Pr|Bt = n, Ut = n) = (0.53, 0.47)$. See also the Hugin network.

Part 2:

The following probabilities can be calculated from the original network by inserting the evidence $Pr = y$ and $Pr = n$, respectively:

$$P(BT|Pr = y) = (0.64, 0.36)$$
$$P(BT|Pr = n) = (0.106, 0.894)$$
$$P(UT|Pr = y) = (0.73, 0.27)$$
$$P(UT|Pr = n) = (0.107, 0.893)$$

Using the calculated probabilities in a nave Bayes structure, we get

$$\begin{aligned}
P(Pr|Bt=n,Ut=n) &= \frac{P(Bt=n,Ut=n|Pr)P(Pr)}{P(Bt=n,Ut=n)} \\
&= \frac{P(Bt=n|Pr)P(Ut=n|Pr)P(Pr)}{P(Bt=n,Ut=n)} \\
&= (0.449, 0.551)
\end{aligned}$$

Note:

- In the second step we exploit that BT and UT are conditionally independent given Pr in the naive Bayes model.
- Dividing with $P(Bt = n, Ut = n)$ simply normalizes the results so that we get a proper conditional probability distribution; we can easily calculate this value based on the numerator in the expression above $P(Bt = n, Ut = n) = P(Bt = n|Pr = y)P(Ut = n|Pr = y)P(Pr = y) + P(Bt = n|Pr = n)P(Ut = n|Pr = n)P(Pr = n)$.

Notice the difference between this result and the result you got from the original network.