Bayesian networks – exercises

Collected by: Jiří Kléma, klema@labe.felk.cvut.cz

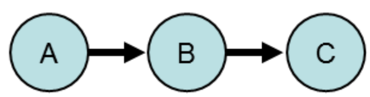
ZS 2012/2013

Goals:

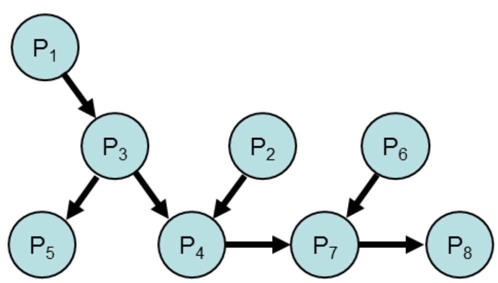
The text provides a pool of exercises to be solved during AE4M33RZN tutorials on graphical probabilistic models. The exercises illustrate topics of conditional independence, learning and inference in Bayesian networks. The identical material with the resolved exercises will be provided after the last Bayesian network tutorial.

1 Independence and conditional independence

Exercise 1. *Formally prove which (conditional) independence relationships are encoded by serial (linear) connection of three random variables.*



Exercise 2. *Having the network/graph shown in figure below, decide on the validity of following statements:*



a) $P_1, P_5 \perp\!\!\!\perp P_6|P_8$,

b) $P_2 \top\!\!\!\top P_6|\oslash$,

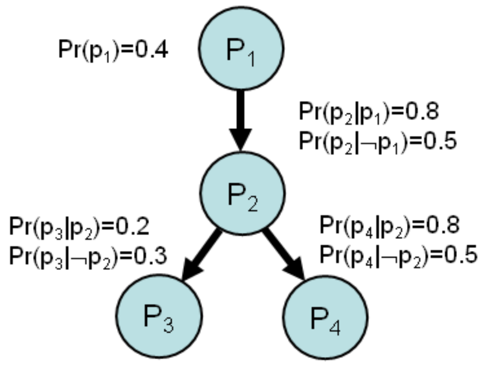
c) $P_1 \perp\!\!\!\perp P_2|P_8$,

d) $P_1 \perp\!\!\!\perp P_2, P_5|P_4$,

e) *Markov equivalence class that contains the shown graph contains exactly three directed graphs.*

2 Inference

Exercise 3. *Given the network below, calculate marginal and conditional probabilities* $Pr(\neg p_3)$, $Pr(p_2|\neg p_3)$, $Pr(p_1|p_2, \neg p_3)$ *a* $Pr(p_1|\neg p_3, p_4)$. *Apply the method of* **inference by enumeration**.



Exercise 4. *For the same network calculate the same marginal and conditional probabilities again. Employ the properties of directed graphical model to* **simplify** *manual computation.*

Exercise 5. *For the same network calculate the conditional probability* $Pr(p_1|p_2, \neg p_3)$ *again. Apply a sampling approximate method. Discuss pros and cons of rejection sampling, likelihood weighting and Gibbs sampling. The table shown below gives an output of a uniform random number generator on the interval (0,1), use the table to generate samples.*

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
0.2551	*0.5060*	*0.6991*	*0.8909*	*0.9593*	*0.5472*	*0.1386*	*0.1493*	*0.2575*	*0.8407*
r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}
0.0827	*0.9060*	*0.7612*	*0.1423*	*0.5888*	*0.6330*	*0.5030*	*0.8003*	*0.0155*	*0.6917*

Exercise 6. *Let us have three tram lines – 6, 22 and 24 – regularly comming to the stop in front of the faculty building. Line 22 operates more frequently than line 24, 24 goes more often than line 6 (the ratio is 5:3:2, it is kept during all the hours of operation). Line 6 uses a single car setting in 9 out of 10 cases during the daytime, in the evening it always has the only car. Line 22 has one car rarely and only in evenings (1 out of 10 tramcars). Line 24 can be short whenever, however, it takes a long setting with 2 cars in 8 out of 10 cases. Albertov is available by line 24, lines 6 and 22 are headed in the direction of IP Pavlova. The line changes appear only when a tram goes to depot (let 24 have its depot in the direction of IP Pavlova, 6 and 22 have their depots in the direction of Albertov). Every tenth tram goes to the depot evenly throughout the operation. The evening regime is from 6pm to 24pm, the daytime regime is from 6am to 6pm.*

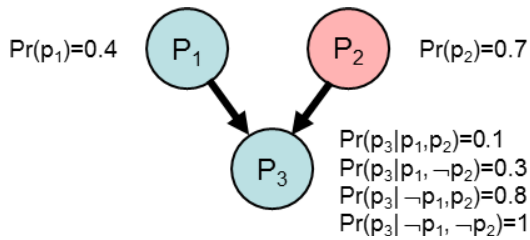
a) Draw a **correct, efficient and causal** *Bayesian network.*

b) Annotate the network with the conditional probability tables.

c) *It is evening. A short tram is approaching the stop. What is the probability it will go to Albertov?*

d) *There is a tram 22 standing in the stop. How many cars does it have?*

Exercise 7. *Trace the algorithm of* **belief propagation** *in the network below knowing that* $\mathbf{e} = \{p_2\}$. *Show the individual steps, be as detailed as possible. Explain in which way the unevidenced converging node* P_3 *blocks the path between nodes* P_1 *and* P_2.



3 (Conditional) independence tests, best network structure

Exercise 8. *Let us concern the frequency table shown below. Decide about independence relationships between A and B.*

	c		$\neg c$	
	b	$\neg b$	b	$\neg b$
a	14	8	25	56
$\neg a$	54	25	7	11

4 Dynamic Bayesian networks

Exercise 9. *A patient has a disease N. Physicians measure the value of a parameter P to see the disease development. The parameter can take one of the following values {low, medium, high}. The value of P is a result of patient's unobservable condition/state S. S can be {good, poor}. The state changes between two consecutive days in one fifth of cases. If the patient is in good condition, the value for P is rather low (having 10 sample measurements, 5 of them are low, 3 medium and 2 high), while if the patient is in poor condition, the value is rather high (having 10 measurements, 3 are low, 3 medium and 4 high). On arrival to the hospital on day 0, the patient's condition was unknown, ie.* $Pr(S_0 = good) = 0.5$.

a) *Draw the transition and sensor model of the dynamic Bayesian network modeling the domain under consideration,*

b) *calculate probability that the patient is in good condition on day 2 given low P values on days 1 and 2,*

c) *can you determine the most likely patient state sequence in days 0, 1 and 2 without any additional computations?, justify.*