Tutorial Bayesian Network WS 2007/08
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Prof. Dr. Dr. Lars Schmidt-Thieme , Leandro Marinho , Krisztian Buza

Tutorial 13 - Solutions

Exercise 1 Types of missingness (20 points)

Show examples for missing data, where

a) [5 pts.] a variable *V* is missing completely at random (MCAR)

No	A	B	C(observed)	C(true value)
1	0	1	0	0
2	0	1	1	1
3	0	1	-	0
4	1	1	-	0
5	1	0	1	1
6	1	0	0	0
7	0	1	-	1
8	0	1	-	1

I(M_C , C(true value))
The probability of missing is independent from the true value of C.
$P(M_C) = P(M_C \mid C(\text{true value}))$

i.e.: P(*C is missing*) = 4 / 8
P(*C is missing* | C(true value)=0) = 2 / 4 (cases 3,4 out of 1,3,4,6)
P(*C is missing* | C(true value)=1) = 2 / 4 (cases 7,8 out of 2,5,7,8)
2 / 4 = 2 / 4 = 4 / 8

b) [5 pts.] a variable *V* is missing at random (MAR),

No	A	B	C(observed)	C(true value)
1	0	1	0	0
2	0	1	1	1
3	0	1	0	0
4	1	1	-	0
5	1	0	1	1
6	1	0	0	0
7	0	1	1	1
8	0	1	-	1

I(M_C , C(true value) | B)
The probability of missing is conditionally independent from the true value of C given B:
$P(M_C \mid C(\text{true value}), B) = P(M_C \mid B)$

i.e.: P (*C is missing* | C(true value) = 1, B = 1) = 1 / 3
B=1 & C(true value)=1 in cases: 2,7,8, among these cases C is missing once: 8
P (*C is missing* | C(true value) = 0, B = 1) = 1 / 3
B=1 & C(true value)=0 in cases: 1,3,4, among these cases C is missing once: 4
P (*C is missing* | B = 1) = 2 / 6
B=1 in cases: 1,2,3,4,7,8 among these cases C is missing twice: 4,8
1 / 3 = 1 / 3 = 2 / 6

and

P (*C is missing* | C(true value) = 1, B = 0) = 0 / 1 = 0
B=0 & C(true value)=1 in case 5, C is not missing in this case
P (*C is missing* | C(true value) = 0, B = 0) = 0 / 1 = 0
B=0 & C(true value)=0 in case 6, C is not missing in this case
P (*C is missing* | B = 0) = 0 / 2 = 0
B=0 in cases: 5,6 among these cases C is never missing.

c) [5 pts.] a variable *V* is missing systematically

No	A	B	C(observed)	C(true value)
1	0	1	-	0
2	0	1	1	1
3	0	1	0	0
4	1	1	-	0
5	1	0	1	1
6	1	0	-	0
7	0	1	1	1
8	0	1	-	1

The probability of missing depends on the true value of C: if the true value of C is 0, it is missing with the probability of 3 / 4, but if the true value of C is 1, it is missing with the probability of 1 / 4.

d) [5 pts.] a variable *V* is hidden.

No	A	B	C(observed)	C(true value)
1	0	1	-	0
2	0	1	-	1
3	0	1	-	0
4	1	1	-	0
5	1	0	-	1
6	1	0	-	0
7	0	1	-	1
8	0	1	-	1

C is hidden (it is missing everywhere).

Exercise 2 Types of missingness: MCAR vs. MAR (8 points)

The following statements are given:
(1) If a set of variables *V* in a dataset *D* is MCAR, then *V* is also MAR in the same dataset
(2) If a set of variables *V* in a dataset *D* is MAR, then *V* is also MCAR in the same dataset
Which of the statements are true? Justify your answer!

Statement (1) is true, because independence is “stronger” then conditional independence (i.e. if the variables X and Y are independent, they are also conditionally independent given any set of the other variables).

Exercise 3 Distribution of completion (7 points)

Suppose we are given the following instance: (*A*=1, *B*=2, *C*=missing, *D*=missing).
The distribution of completion should be the uniform distribution. Fill up the missing cells of the probability distribution table above.

A	*B*	*C*	*D*	*p*(*A*,*B*,*C*,*D*)
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	0
1	0	0	0	¼
1	0	0	1	¼
1	0	1	0	¼
1	0	1	1	¼
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	1	0