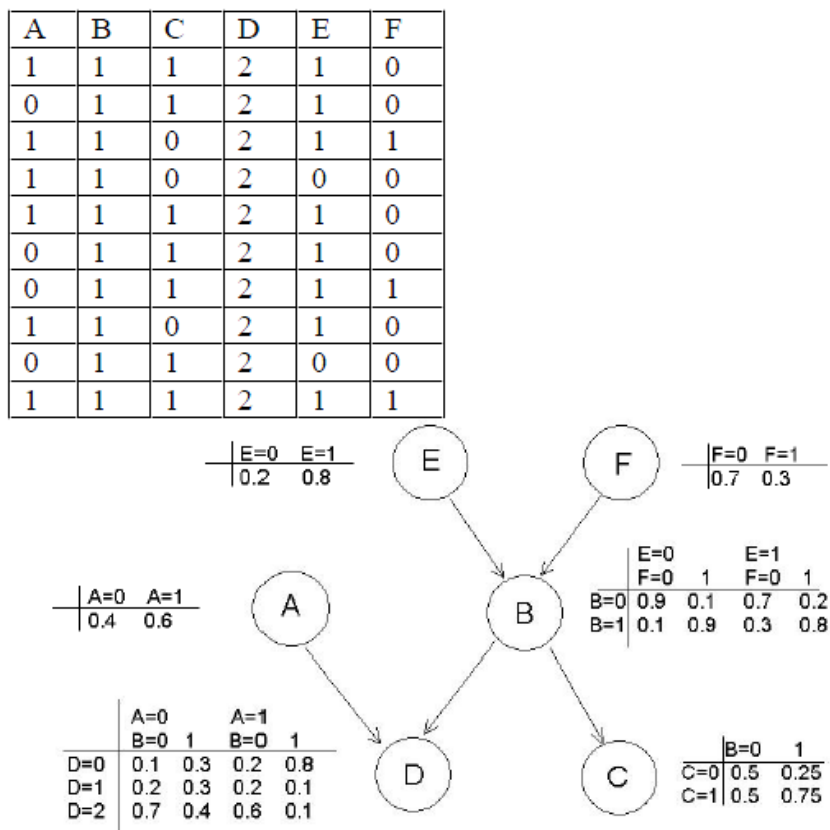# Übung 7

Lösungen bitte via Moodle / learnweb, bis zum 06.07.2010 einreichen.

Aufgabe 1 Importance sampling   (20 Points)

Suppose we apply (1) self importance sampling (2) adaptive sampling on the following Bayesian network conditioned by the evidence $Ev = \{D=2, B=1\}$.
In the first step we generated the sample instances above.

A	B	C	D	E	F
1	1	1	2	1	0
0	1	1	2	1	0
1	1	0	2	1	1
1	1	0	2	0	0
1	1	1	2	1	0
0	1	1	2	1	0
0	1	1	2	1	1
1	1	0	2	1	0
0	1	1	2	0	0
1	1	1	2	1	1

	E=0	E=1
	0.2	0.8

	F=0	F=1
	0.7	0.3

	A=0	A=1
	0.4	0.6

	E=0 F=0	E=1 F=0
B=0	0.9 0.1	0.7 0.2
B=1	0.1 0.9	0.3 0.8

	A=0 B=0 1	A=1 B=0 1
D=0	0.1 0.3	0.2 0.8
D=1	0.2 0.3	0.2 0.1
D=2	0.7 0.4	0.6 0.1

	B=0 1
C=0	0.5 0.25
C=1	0.5 0.75

a) [5 pts.] Which conditional probability distributions will be changed?
b) [5 pts.] What are the new conditional probability distributions for the second step in case of self importance sampling?
c) [5 pts.] What are the new conditional probability distributions for the second step in case of adaptive sampling?
d) [5 pts.] What is the most important difference between self importance samling and adaptive sampling (based on slide 4 in bayes-10-approxinference-propagation.pdf)?

Aufgabe 2   Basics of parameter learning   (10 Points + 5 Bonus)

a) [5 pts.] Suppose, we observe the following values of a discrete probabilistic variable V: 0, 3, 0, 1, 2, 1, 1, 2, 2, 2, 2, 1, 0, 1, 2, 2, 1, 0, 2, 1, 3, 3, 1, 2, 2, 1, 1, 3. Which distribution can be learned via maximum likelihood principle?

b) [5 pts.] Suppose we are interested in families with 3 children. Suppose V denotes the amount of daughters in the family. What is the "theoretical" distribution of V?

c) [Bonus 5 pts.] Which distribution is the theoretical distribution of V "similar" to? (Hint: $V = V_1 + V_2 + V_3$ where: $V_1 = 1$ if the first child is a girl, $V_1 = 0$ else. $V_2 = 1$ if the second child is a girl, $V_2 = 0$ else. $V_3 = 1$ if the third child is a girl, $V_3 = 0$ else.)

Aufgabe 3 Types of missingness   (10 Points)

Show examples for missing data, where
a) [2,5 pts.] a variable V is missing completely at random (MCAR),
b) [2,5 pts.] a variable V is missing at random (MAR),
c) [2,5 pts.] a variable V is missing systematically,
d) [2,5 pts.] a variable V is hidden.

The example should not be the same as the example in the lecture notes.

Aufgabe 4 Types of missingness: MCAR vs. MAR   (5 Bonus)

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!