Algorithmical and Statistical Modeling – Fall 2010 – Final Exam

Your name:

Problem 1 (20 pts). Consider the following situation concerning animal population monitoring. Let the map of Germany be partitioned into 10 by 10 km squares S_{ij}, where the index pairs (i, j) come from some index set I and are chosen such that S_{ij+1} is the square east of S_{ij}, and S_{i+1j} is the square south of S_{ij}. Assume that since 1950, every year k (k = 1950, ..., 2010) up to 2010, in Spring in each of the squares S_{ij} the number $N(i, j, k)$ of male blackbirds has been determined by amateur ornithologists (which in fact is what is done). Assume that there are no missing values, that is, every $N(i, j, k)$ has actually been determined.

Based on these data, a biology PhD student with some background in statistical modeling wants to estimate a probabilistic model which should provide a succinct account of (i) the spatial interaction of blackbird populations, i.e. some account of how the population in one square interacts with the population in the eight neighboring squares, and (ii) the temporal dynamics of blackbird populations over the years.

(a, 4 pts) Outline a reasonable probability model by

1. specifying measure space(s) and random variables;
2. describing in plain English Ω and ω from the underlying probability space $(\Omega, \mathcal{A}, P)$.

(b, 4 pts) Specify how you would arrange your random variables in a Bayesian network, having the PhD students research objectives in mind (delivery format: either a graphical sketch or a formal specification).

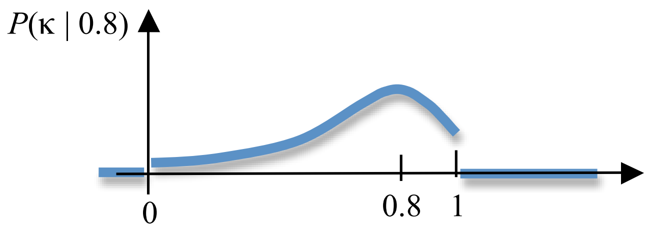
(c, 4 pts) Explain what modeling assumptions concerning blackbird population dynamics went into your Bayesian network.

(d, 4 pts) Specify how you would represent a probability distribution over your chosen random variables.

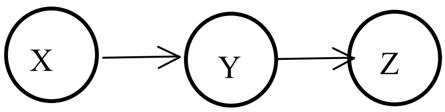
(e, 4 pts) Specify how you would estimate the parameters in your distribution representation from the observed data.

Problem 2 (20 pts). Imagine you were a physicist and you had discovered the existence of a new natural constant κ. The theory behind your discovery requires that κ lies between 0 and 1, but does not constrain it further. The true value has thus to be determined by measurement. You build a measurement apparatus whose measurement error you know is normal distributed, i.e. it will return measurements which are $\mathcal{N}(\kappa, 1)$ distributed. Using this apparatus is extremely expensive, so your budget allows only a single measurement. It turns out to give a value of 0.8. Based on this finding you publish a paper. What would you claim in this paper about the true value of κ, and what would be the reasoning behind your claim? (a formula plus a graphical sketch will be helpful).

Solution. Using a Bayesian approach, the prior on κ is the uniform distribution over the unit interval. Plotting $P(\kappa \mid 0.8) = P(0.8 \mid \kappa)\, P(\kappa) / P(\text{measurement}) \sim P(0.8 \mid \kappa)$ $P(\kappa)$ gives sth. like this, with the BMP estimate $\int x\, P(\kappa \mid 0.8)\, dx$ of κ lying somewhat below 0.8.



Problem 3 (20 pts). Here is a simple Bayesian network:



All RV's are binary, and the probability tables are

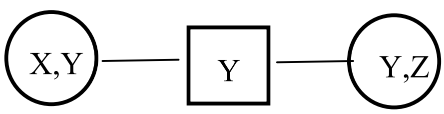
P(X= 0)	P(X = 1)
0.5	0.5

X	P(Y= 0|X)	P(Y = 1|X)
0	0.5	0.5
1	0.25	0.75

Y	P(Z= 0|Y)	P(Z = 1|Y)
0	0.5	0.5
1	0.75	0.25

Your task: transform this Bayesian network into a join tree, including the calculation of a set of belief potentials.

Solution. The obvious (unique) join tree structure is this:



Let p, q, r denote the belief potentials of X,Y – Y -- YZ, respectively.

We follow the recipe from the lecture notes. Here is a synopsis of the operations:

Initialization: