Final Exam, Algorithmical and Statistical Modelling, Fall 2007 – Solutions[1]

Note. The points of problems on this sheet sum to 110, of which a maximum of 100 will be counted toward the course grade.

Problem 1 (20 points). A manufacturer of memory chips claims, *"the probability of a bit error in our SuperMem memory chips is 1 in 10^{15} access operations"*. Your task: formalize this claim as a probability statement, introducing an underlying event space Ω and a suitable random variable (or suitable random variables if you think several are needed). **Deliverables:** (i) explain Ω and your random variable(s) in words, and (ii) give a mathematical expression which captures the manufactuer's statement.

Solution. One (I would say natural) modelling approach is to view the access operations on a given chip in a similar way as a sequence of i.i.d. experiments carried out in a particular Lab (as in the slides of our extra session on basics of statistical modelling). This leads to $\omega \in \Omega$ being interpreted as "the sequence of access operations that are effected at a particular chip from the SuperMem series". The variables X_i (where i = 1, 2, 3, …) can be designed to take values in {0, 1}, where $X_i(\omega) = 1$ iff the i-th access to that particular chip produces a bit error, and 0 if it doesn't. Assuming that the X_i are i.i.d. (a questionable assumption, first because a chip ages and second, because if there is a bit error at some time, it is likely that it will soon be repeated if it is caused by some damage), one can identify all of these X_i with a generic RV X having the same distribution as any of the X_i. The formal statement is then $P(X = 1) = 10^{-15}$, or equivalently, because our X is an indicator variable, $E[X] = 10^{-15}$, or equivalently, $P_X(1) = 10^{-15}$.

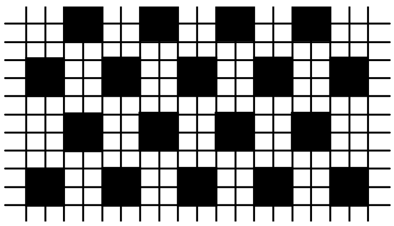
Problem 2 (30 points) Consider the Markov chain $(X_n)_{n = 0,1,2...}$ with values taken in the integers $\mathbb{Z}$, specified by (i) $P(X_0 = 0) = 1$; (ii) $P(X_{n+1} = X_n) = 1/2$; (iii) $P(X_{n+1} = X_n+1) = P(X_{n+1} = X_n-1) = 1/4$.
Show that for all n, $E[X_{n+1}^2] = E[X_n^2] + 1/2$. *Hint:* this can be done by a straightforward computation, which at one point needs a little "aha" idea to find the right transformation of a formula.

Solution. Proof by direct computation. The "aha" takes you from the 3rd to the 4th line.

$$
\begin{aligned}
E[X_{n+1}^2] &= \sum_{z \in \text{Integers}} P(X_{n+1} = z) z^2 = 1 \\
&= \sum_z \left(\frac{1}{2} P(X_n = z) + \frac{1}{4} P(X_n = z-1) + \frac{1}{4} P(X_n = z+1) \right) z^2 \\
&= \frac{1}{2} E[X_n^2] + \frac{1}{4} \sum_z P(X_n = z-1) z^2 + \frac{1}{4} \sum_z P(X_n = z+1) z^2 \\
&= \frac{1}{2} E[X_n^2] + \frac{1}{4} \sum_z P(X_n = z)(z+1)^2 + \frac{1}{4} \sum_z P(X_n = z)(z-1)^2 \\
&= \frac{1}{2} E[X_n^2] + \frac{1}{4} \sum_z P(X_n = z)(z^2 + 2z + 1) + \frac{1}{4} \sum_z P(X_n = z)(z^2 - 2z + 1) \\
&= \frac{1}{2} E[X_n^2] + \frac{1}{4} \sum_z P(X_n = z)(2z^2 + 2) = E[X_n^2] + \sum_z P(X_n = z) \frac{1}{2} = \\
&= E[X_n^2] + \frac{1}{2}.
\end{aligned}
$$

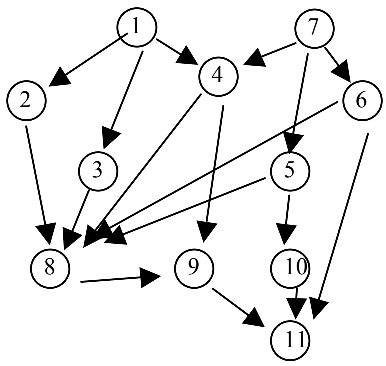
[1] Second version. The model solution in the first version had a wrong "solution" for problem 4.

Problem 3 (20 points) At the bottom (right) is a picture by Pop Art artist Roy Lichtenstein, titled "HIM". Graphit and touche on paper, 22 x 17 inches, 1964. Saint Louis Art Museum. Taken from http://www.lichtenstein-foundation.org). Lichtenstein's hallmark is his imitation of raster points (as known from printing) in his paintings – the dots you see on this image are hand-drawn... Your task: design a "Lichtenstein dot detector" in the form of Markov random field (MRF), which gets black and white images like this HIM picture as input (= activation pattern on visible units, which correspond to pixels) and develops, through its stochastic update dynamics, a segmentation indicator for the "Lichtenstein-dotted" ares. More precisely, over image areas with Lichtenstein dots, the hidden MRF layer units should develop an activation of +1, while in non-dotted areas, their activation should develop toward –1. – Notice that the Lichtenstein dots are larger than the image pixels, that is, one Lichtenstein dot corresponds to a small cluster of image pixels. Let's say, for simplicity, that each Lichtenstein dot covers 2 x 2 image pixels, with blank pixels in between according to the pattern shown in the schema below (we assume that the Lichtenstein dots all have the same size across different Lichtenstein paintings – I hope you don't mind that we pretend, for this exam, that the world is simple). **Deliverables:** A description of the MRF topology and RVs, plus an energy function for the MRF, plus an explanation in words of the energy component(s) that you choose. You may assume that a local pattern matching algorithm P is given, which gets a 5x5 binary pixel image as input and returns 1 if the input image corresponds to some 5x5 subsection of a Lichtenstein dot pattern, and returns 0 else.



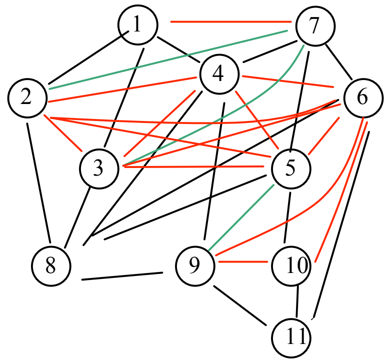
Solution. (One possibility – there is some freedom of design). As in the MRF example of the lecture notes, we set up the topology of the MRF by creating two rectangular grids of same (= image pixel) size, of visible (y_i) and hidden (x_i) nodes. The energy function has two types of components. The first component encourages coherent areas of all +1's or all –1's to emerge in the hidden units x_i, and (like in the "Bayes networks" pattern denoising example in the lecture notes) is coded by a component $-\alpha \sum_{i,j} x_i x_j$, where the sum is taken over all pairs i, j of neighboring cells. The second component implements the following intuition: "for every pixel index i, decrease the energy iff (i) the pixel neighborhood of i in the visible layer matches the Lichtenstein dot pattern and (ii) the hidden unit x_i has a value of +1; else – increase the energy. Let p_i be the 5x5 input (visible) pattern centered on pixel i. Then, the second component of our energy is $-\beta \sum_i x_i P(p_i)$, which altogether gives an energy function $E(\mathbf{x}, \mathbf{y}) = -\alpha \sum_{i,j} x_i x_i - \beta \sum_i x_i P(p_i)$. – This is a frugal solution which could be refined in many ways, e.g. accounting for special conditions at image borders or Lichtenstein dot area borders, or using a graded pattern matching P that tolerates some noise in the pattern matching.

Problem 4 (15 points). To the right you see a structural sketch of a BN. Derive a join tree from it. **Deliverables:** (i) a graphical representation of the triangulated UGM which you get as an intermediate step, (ii) a graphical representation of the join tree that you obtain. In the join tree representation, you may omit the sepsets. (Be aware of the recent correction in the lecture notes: join tree clusters correspond to *maximal* cliques)

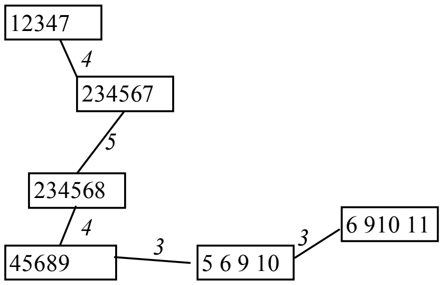


Solution.

(i)



ii)



Notes. The red lines in (i) are the connections that are added by moralizing, the green ones are added by triangulation. This solution was found by Kevin Murphy's BN toolbox. When I posed the question and did the first solution manually (which I posted in the original model solution), I only saw one of the three triangulation necessities (the link 9-5 which triangulates the loop 4 5 10 9). The other two are needed to fill the loop 1 7 5 3 2, which I did not see when I worked out my solution by hand. This omission yielded wrong clusters, which in turn led to a join tree which violated the running intersection property (in the first posting of the model solution), as was pointed out to my be Vytenis. It seems that this graph complexity exceeds what humans can overview in a few minutes… So in the grading of this question, I will not expect perfect solutions, and give full score whenever the basic principle seems appreciated and followed, and the provided graphics are clearly drawn.

Problem 5 (25 points). Write a short essay (target size 200 – 300 words, about as much or slightly less as the text for problem 3 above) on the topic "Do humans use fuzzy control when they try to control themselves, or something in their environment?" Your essay will be graded according to insight, clarity, wit.

Solution. *(I am curious to see what you'll write… the idea behind this problem being that human control actions are very diverse, ranging from millisecond automatisms (like the eye blink reflex) to lifespan optimal control (e.g., choosing an education to pave the path to a desired carreer); the explicit-verbal nature of fuzzy control being an appropriate model only for some cases, and for others, clearly not.)*