CSCE 582 Spring 2012

Final Exam

2012-04-28

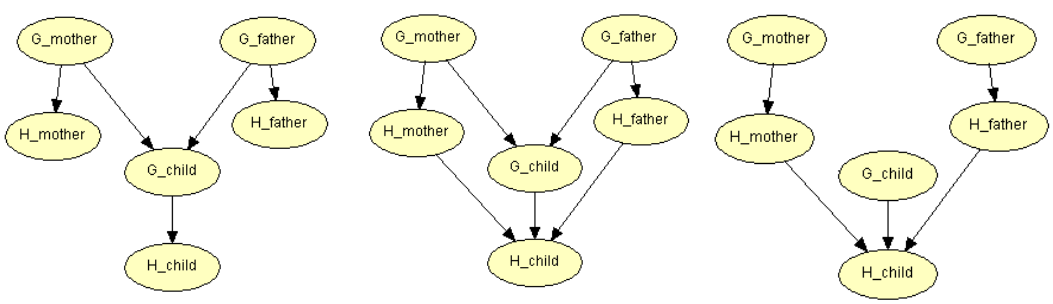
Closed book and notes

Question 1 (60 points)

This question is from Chapter 14 of: Russell, Stuart and Norvig, Peter. *Artificial Intelligence: A Modern Approach, 3rd edition*, Prentice-Hall, 2010.

Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness in inherited by a simple mechanism: that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out to be the mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of random mutation flipping the handedness.

a. Which of the three networks below claim that $P(G_{father}, G_{mother}, G_{child}) = P(G_{father})P(G_{mother})P(G_{child})$?



(a) (b) (c)

Answer: (c): there is a diverging connection between each pair in G_{father}, G_{mother}, G_{child}.

b. Which of the three networks make independence claims that are consistent with the hypothesis about the inheritance of handedness?

   Answer: (a) and (b). Notes: (b) can mimic (a); (c) imposes the constraint that the gene of the child is (unconditionally) independent of the genes of mother and father, which is contrary to the hypothesis.

c. Which of the three networks is the best description of the hypothesis?

Anwer: (a). Note: (b) can mimic (a), but it is not faithful to the hypothesis, because it allows for additional dependencies, such as that the handedness of the child depends directly (and not through genes) on the handedness of the parents.

d. Write down the CPT for the G_{child} node in network (a) in terms of s and m.

G_{father}	*l*	*l*	*r*	*r*
G_{mother}	*l*	*r*	*l*	*r*
l	*1-m*	*1/2*	*1/2*	*m*
r	*m*	*1/2*	*1/2*	*1-m*

e. Suppose that P(G_{father} = l) = P(G_{mother}=l) = q. In network (a), derive an expression for P(G_{child} = l) in terms of m and q only, by conditioning on its parent nodes. Start as follows:

$$P(G_{child} = l) = \sum_{g_m, g_f} P\big(G_{child} = l \big| g_m, g_f\big) P\big(g_m, g_f\big) = \cdots$$

Answer:

$$\begin{aligned} P(G_{child}=l) &= \sum_{g_m,g_f} P(G_{child}=l|g_m,g_f)P(g_m,g_f) \\ &= \sum_{g_m,g_f} P(G_{child}=l|g_m,g_f)P(g_m)P(g_f) \\ &= (1-m)q^2 + 0.5q(1-q) + 0.5(1-q)q + m(1-q)^2 \\ &= q^2 - mq^2 + q - q^2 + m - 2mq + mq^2 \\ &= q + m - 2mq \end{aligned}$$

f. Under conditions of genetic equilibrium (i.e., the prior probability of left-handedness is the same for everyone), we expect the distribution of genes to be the same across generations. Use this to calculate the value of q, and, given what you know about handedness in humans, explain why the hypothesis described at the beginning of the question must be wrong.

Answer:

f. Equilibrium means that $P(G_{child} = l)$ (the prior, with no parent information) must equal $P(G_{mother} = l)$ and $P(G_{father} = l)$, i.e.,

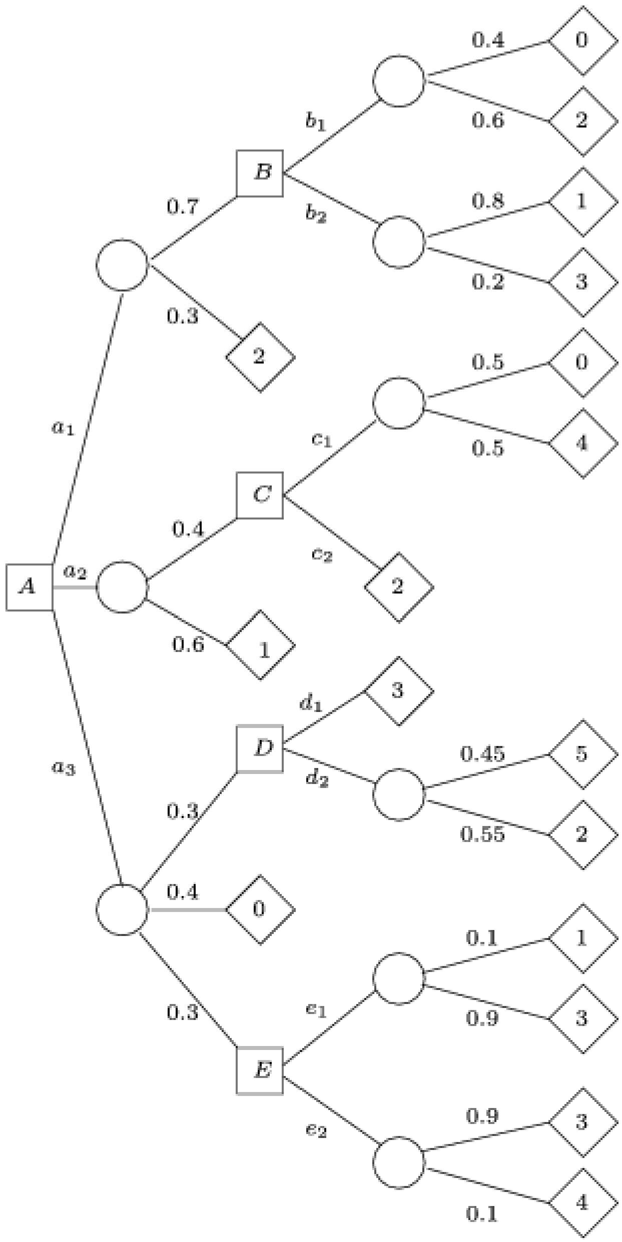
$$q + m - 2mq = q, \text{ hence } q = 0.5.$$

But few humans are left-handed ($x \approx 0.08$ in fact), so something is wrong with the symmetric model of inheritance and/or manifestation. The "high-school" explanation is that the "right-hand gene is dominant," i.e., preferentially inherited, but current studies suggest also that handedness is not the result of a single gene and may also involve cultural factors. An entire journal (*Laterality*) is devoted to this topic.

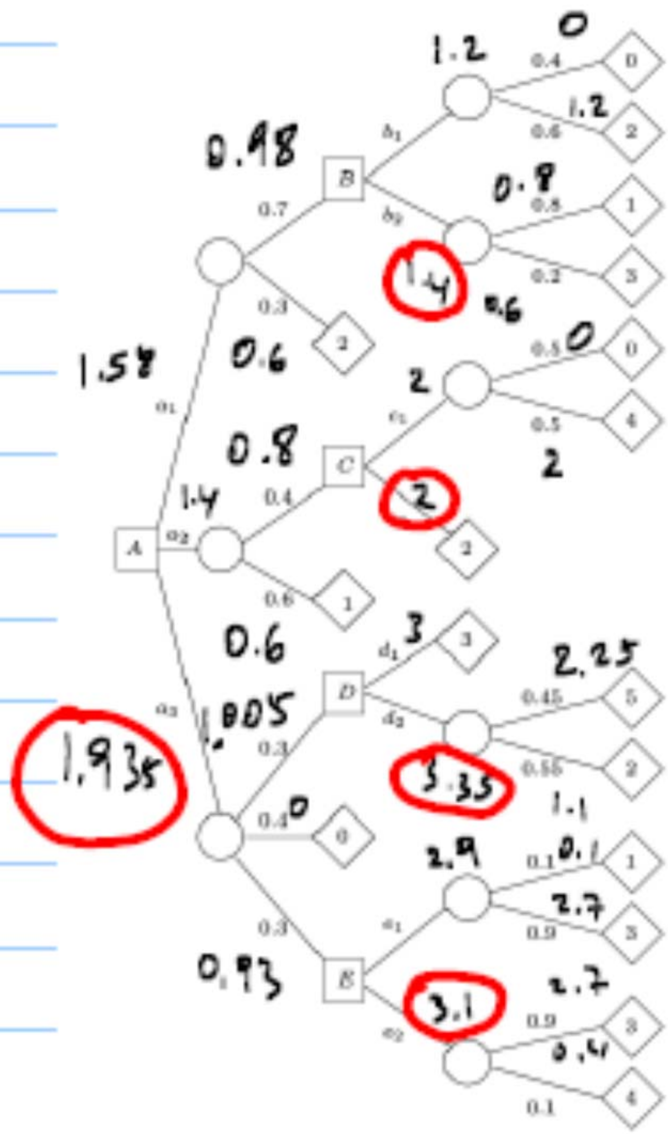
Question 2 (30 points)

This is exercise 9.8 in the textbook.

Solve the decision tree in the figure below.



Answer: (Jensen & Nielsen) The optimal strategy is to choose a_3. If the observation is the upper branch, then choose d_2, if lower branch, then choose e_2. The expected utility of the strategy is 1.905. My computation gives 1.935 instead of 1.905---see below.



A
B
C
D
E
0.7
0.3
0.4
0.6
0.3
0.4
0.3
0.4
0.6
0.8
0.2
0.5
0.5
0.45
0.55
0.1
0.9
0.9
0.1
1.2
0
1.2
0.98
0.8
1.4
0.6
1.58
0.6
0
0.8
2
2
1.4
2
0.6
3
2.25
1.005
1.935
3.35
1.1
0
2.9
0.1
2.7
0.93
3.1
2.7
0.4

(tie)

Question 3 (30 points)

This question is based on: Allan L. Jensen and Finn V. Jensen. "MIDAS---An Influence Diagram for Management of Mildew in Winter Wheat." *Proceedings of UAI-96*, pp.349-356. Also see http://www.hugin.com/developer/samples/mildew.

a. Approximately two months before harvesting a wheat field a farmer can observe the state of the crop and can observe whether it has been attacked by mildew. If there is an attack, the farmer should decide on a treatment with fungicides. Draw an influence diagram with the following nodes:

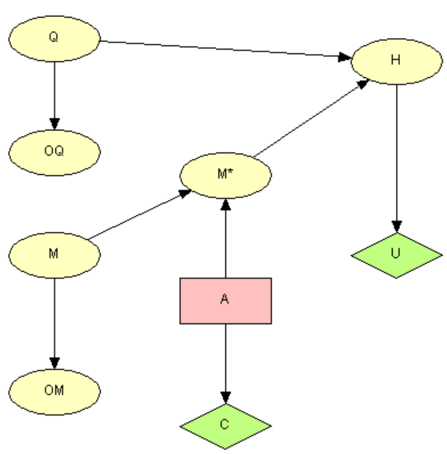
Six chance nodes:

- *The current state of the crop, **Q** with states fair (f), average (a), good (g) and very good (v)*
- *The current mildew-situation, **M** with states no, little (l), moderate (m) and severe (s)*
- *The mildew situation after fungicide-treatment, **M***, which has the same states as **M**, and which has the same distribution as **M** if there is no fungicide-treatment.*
- *The state of the crop at harvest time, **H** with the states from **Q** plus rotten (r), bad (b) and poor (p)*
- *The observation, **OQ**, of **Q***
- *The observation, **OM**, of **M***

*One action node, **A** (modeling the fungicide-treatment) with actions no, light (l), moderate (m) and heavy (h).*

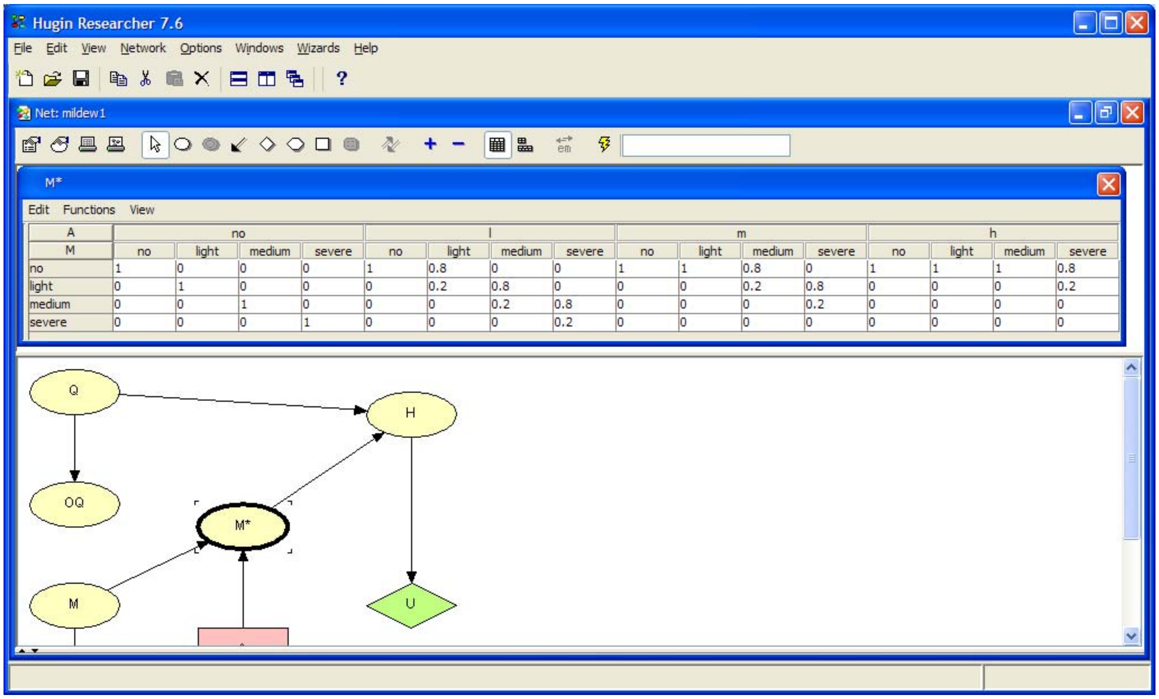
*Two utility nodes: **C** (the cost of treatment) and **U** (the price obtained at harvest time).*

Answer: (mildew1.net from Samples in Hugin distribution)



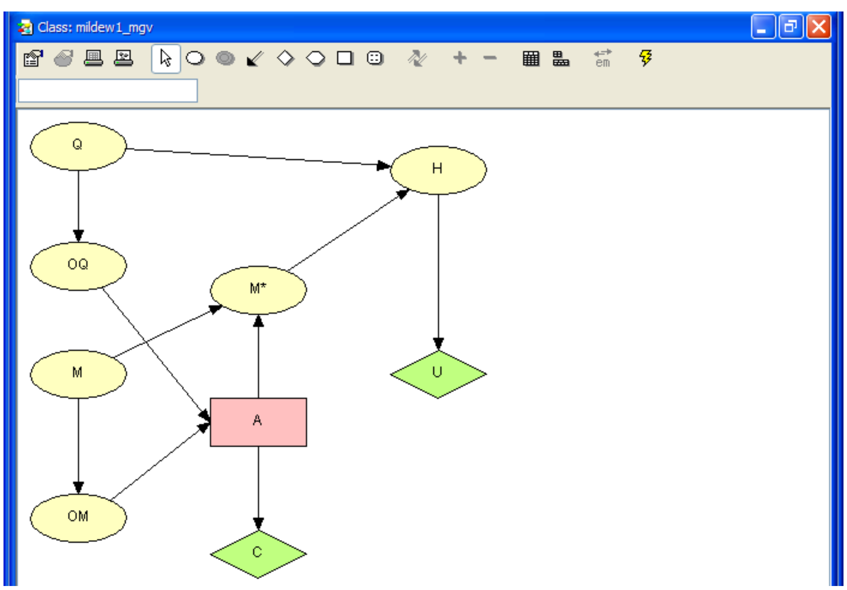
b. Draw the CPT for the node M*, with a reasonable choice of numbers. (There is a single correct solution for some of the entries, while for other entries you should simply guess.)

Answer (from Hugin distribution): Note that the entries for “no treatment” are fixed.



A	no				l				m				h			
M	no	light	medium	severe	no	light	medium	severe	no	light	medium	severe	no	light	medium	severe
no	1	0	0	0	1	0.8	0	0	1	1	0.8	0	1	1	1	0.8
light	0	1	0	0	0	0.2	0.8	0	0	0	0.2	0.8	0	0	0	0.2
medium	0	0	1	0	0	0	0.2	0.8	0	0	0	0.2	0	0	0	0
severe	0	0	0	1	0	0	0	0.2	0	0	0	0	0	0	0	0

Alternative solution for A is given below:

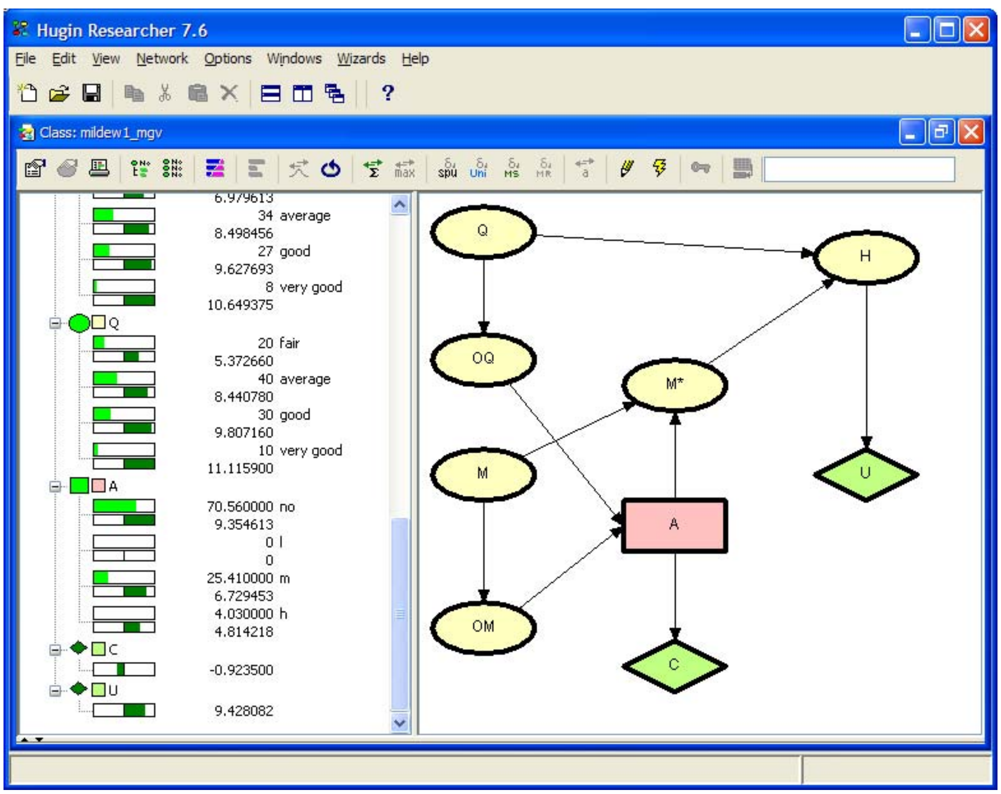


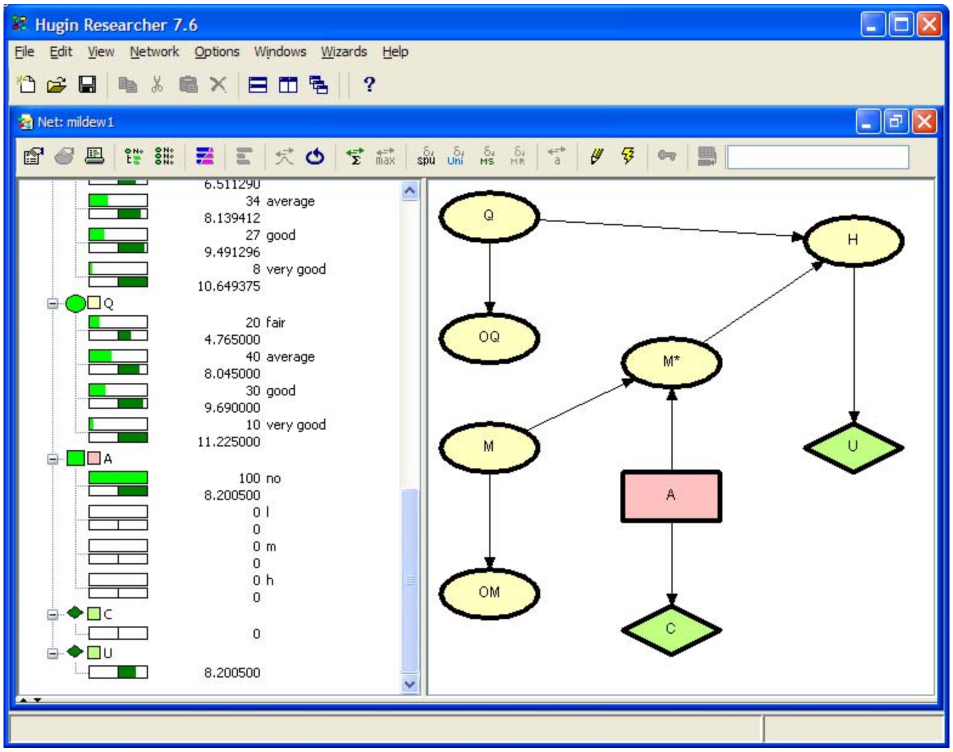
Note that the results for this network are the same as those from the network provided in the Hugin distribution (mildew1.net), at least for several configurations tested:

1. When OM= Severe and OQ = very good, mgv gives A=no (EU=8.389), original gives same.
2. When OM= no and OQ= fair, original gives A=no (EU = 8.4859), mgv gives same.
3. When OM= medium and OQ = good, mgv gives A= m (EU = 7.962), original gives same.

However:

4. When no evidence, original gives A=no (EU=8.2005), mgv gives A=no (EU=9.3546). I think that mgv refuses to give an optimal policy in the absence of observations, because the extra links indicate that OQ and OM are always observed before A is decided. Note that there is a distribution for A in mgv, while there is an optimal policy for original:




Hugin Researcher 7.6
File Edit View Network Options Windows Wizards Help
Net: mildew1
6.511290
34 average
8.139412
27 good
9.491296
8 very good
10.649375
Q
20 fair
4.765000
40 average
8.045000
30 good
9.690000
10 very good
11.225000
A
100 no
8.200500
0 l
0
0 m
0
0 h
0
C
0
U
8.200500
Q
H
OQ
M*
M
U
A
OM
C