NATIONAL UNIVERSITY OF SINGAPORE
SCHOOL OF COMPUTING
SOLUTIONS FOR FINAL EXAMINATION FOR
Semester 2, 2007/2008

CS3243 - Foundations of Artificial Intelligence

2 May 2008 Time Allowed: 2 Hours

INSTRUCTIONS TO CANDIDATES

1. The examination paper contains **FOUR (4) questions** questions and comprises **TWELVE (12) pages**.
2. Weightage of questions in given in square brackets. The maximum attainable score is 100.
3. You are allowed to bring **TWO** double-sided A4 sheets of notes for this exam.
4. You are allowed to use a calculator in this exam, but laptops are not allowed.
5. Write all your answers in the space provided in this booklet.
6. **Please write your matriculation number below.**

MATRICULATION NUMBER: ____________________

(this portion is for the examiner's use only)

Question	Marks	Remark
Q1		
Q2		
Q3		
Q4		
Total		

Question 1 : Short Structured Questions [20 marks]

CS3243 is a survey course of AI techniques. As a warm-up exercise for this exam, the following are short structured questions to test your understanding of some basic concepts that were introduced in CS3243.

A. Suppose Blizzard announces that they have made a great technological breakthrough – the *World of Warcraft* now has NPC (Non-Player Characters) that are indistinguishable from regular player characters, i.e. it is impossible to tell whether a given character in the game is controlled by a human player or a computer. Has Blizzard succeeded in writing a program that passes the *Turing Test*? Explain. [3 marks]

It depends on what Blizzard means by "indistinguishable from regular player characters." The Turing test consists of a human judge engaging in conversation with a human and an AI agent and the agent is said to pass the test if the human judge cannot distinguish between the two. This situation can most certainly be replicated in a MMORPG like World of Warcraft, so the question is whether the NPC agents will interact with the human judge in conversation or if only in-game actions are involved, i.e. the NPCs fight like human players. Credit will be given if the student demonstrates an understanding of what the Turing test is about and supports his view with reasonable arguments.

The crux of this matter is that the students should have an appreciation that the context of the interaction is important and when someone claims "it is impossible to tell whether a given character in the game is controlled by a human player or a computer," what does he really mean.

B. Many problems can be solved with both informed search and local search. Give an example where informed search is preferred to local search and another example where local search would be preferred. Explain. [4 marks]

Credit will be given for any reasonable answers. E.g. informed search may be preferred in problems for which local search often ends up in a local minimum or if it is not easy to find a good heuristic function for local search. Informed search is also preferred if the path cost matters.

Local search is preferred in problems for which path cost does not matter and local search is often faster, i.e. CSPs, WALKSAT versus DPLL.

C. Give a PEAS analysis for *a recommendation system for books* in the context of an online shopping portal like Amazon.com. The system provides users with a list of recommended books (to buy) and collects data on the books purchased by users and also the ratings on books owned by the users. Full credit will only be given for answers that are fully justified by written comments.

Performance Measure [2 marks]: *Credit will be given for any reasonable answer, but most obvious performance measure is sales, i.e. the probability that a customer will buy a recommended book.*

Environment [3 marks]

The following is a list of likely answers. Other answers would be acceptable if well-substatiated by the student.

- Observability: *Fully Observable*

- Determinism: *Deterministic*
- Episodic/Sequential: *Sequential*
- Static/Dynamic: *Static*
- Discrete/Continuous: *Discrete*
- Single/Multi-agent: *Single*

Sensors [2 marks]: *User ratings & user purchases obtain through web form.*
Actuators [2 marks]: *Web page display with recommendations.*

D. Alice wrote an agent for Breakthrough that employs Minimax. Unfortunately, the agent seems to perform relatively badly compared to the other agents written by the students of CS3243. Bob suggests that Alice should try to improve the performance of the agent by adding alpha-beta pruning. Do you agree with Bob? Explain. [4 marks]

It depends on how the student interprets the problem. What does it mean by perform badly?

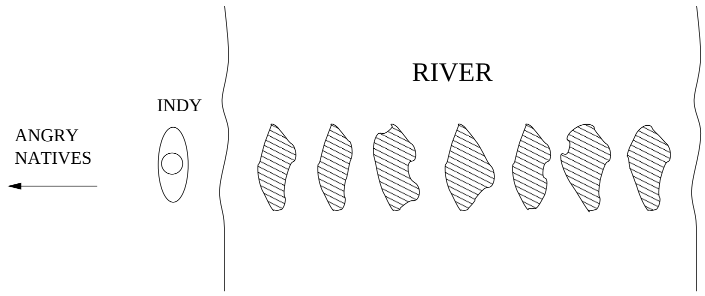
If time/depth is the limiting factor, then alpha-beta will improve performance since it is perhaps possible to increase the depth of the search.

However, the problem might lie in the evaluation function. If the evaluation function is really bad, then alpha-beta might not improve the agent's performance at all. It is also important to note that alpha-beta does not actually change the outcome, only improves the speed.

The crux of this problem is that the student must demonstrate that he/she understands that alpha beta doesn't necessarily improve performance.

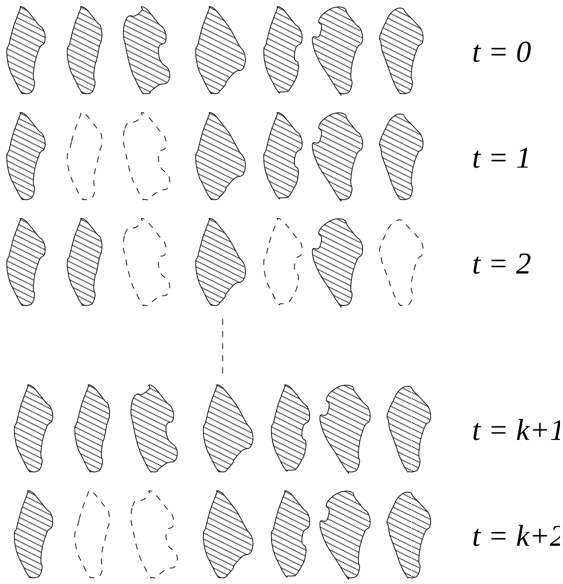
Question 2 : Indiana Jones and the River of Death [16 marks]

Indiana Jones has just recovered another lost artifact. Indy is standing on the western bank of a river and he has a group of angry natives in hot pursuit. Across the river lies a series of 7 stones, arranged in a straight line, and each stone is 1 meter away from its immediate neighbours or the two river banks.



Let us divide the time into intervals of one second each. At the start of each interval, Indy can hop once from a stone/river bank to any stone/bank that is not more than 1.5 meters away, or stay put. He can hop so fast that we assume the time taken per hop is negligible.

The tricky part is this: The stones may sink and resurface! Within a time interval, a stone may sink at the middle of the interval, remains submerged and may resurface at the middle of another interval. If Indiana Jones is standing on a sinking stone, he will drown. Of course, Indiana Jones cannot hop to a stone that is submerged. At the zeroth interval, all stones are above the water and Indiana Jones is ready to hop. He has already derived the pattern of sinking/resurfacing for each stone during the next few intervals and realizes that the pattern repeats itself after k different configurations. The following figure shows an example of the pattern of sinking and resurfacing stones over time.



Indy has one minute (60 seconds) to cross the river before the angry natives catch up with him. Can you help Indy escape?

A. Formulate this as a search problem. Propose a representation for the state of this problem. State the initial and end states in your representation. [6 marks]

We can represent the state as a tuple (p,t), where p is Indy's position and t is the time elapsed. Note that the configuration of the stones is uniquely specified by $t(\text{mod}\,k)$.

Initial state: $(0,0)$

End state: $(8,T), T \leq 60$

Note: the above is somewhat minimal. Many students included the state of the stones, which is quite irrelevant since the configuration is uniquely determined by $t(\text{mod}\,k)$ as stated above. Full credit is however still given without penalty in such cases.

B. Which <u>uninformed</u> search technique would you apply to help Indy find the fastest way to cross the river? Explain. [4 marks]

Either breadth-first search or iterative deepening will work since we want optimality and each step is of unit cost.

C. Suppose Indy wants to cross the river with the fewest number of hops. Explain how you can formulate the search problem to find a way to cross the river with the <u>fewest number of hops</u> (and yet cross within 60 seconds). [6 marks]

Add the hop count h to the state, i.e. (p,t,h). Instead of using unit cost for each move, the cost for a move will depend on whether p changes. If p changes, the cost is one, otherwise cost will be zero. Instead of using breadth-first or IDS, we will have to use uniform cost search and expand the nodes with minimal hop count first.

No credit is given for students who suggest using A search or only improving the heuristic since this is a DIFFERENT problem and requires a different formulation that requires the hop count to be taking into account for the cost. Time still remains as a factor but is no longer the cost function.*

Question 3 : Validity and Satisfiability [24 marks]

A. What does it mean for a sentence to not be satisfiable? [2 marks]

A sentence is not satisfiable if there does not exist an assignment for which the sentence will evaluate to true.

B. Prove by <u>resolution</u> that the following sentence is not satisfiable:

$$(\forall x \exists y\, p(x,y)) \wedge (\exists y \forall x \neg p(y,x))$$

[10 marks]

This is a test that the student understands the mechanics for resolution and that to prove something is not satisfiable, all we need to do is to resolve to the empty clause.
This sentence consists of two clauses:

$$\forall x \exists y\, p(x,y) \tag{1}$$
$$\exists y \forall x \neg p(y,x) \tag{2}$$

Skolemize and standardize apart:

$$\forall x\, p(x,A(x)) \tag{3}$$
$$\forall z \neg p(B,z) \tag{4}$$

Dropping the universal quantifiers:

$$p(x,A(x)) \tag{5}$$
$$\neg p(B,z) \tag{6}$$

Unify with $\theta = \{x/B, z/A(x)\}$

C. What does it mean for a sentence to be valid? [2 marks]

A sentence is valid if it is always true.

D. Prove by <u>resolution</u> that the following sentence is valid:

$$(\exists y \forall x\, p(x,y)) \Rightarrow (\forall x \exists y\, p(x,y))$$

[10 marks]

This is a test that the student understands the to prove something is valid, we try to resolve **the negation** *to the empty clause.*
To prove that statement is valid, we prove the negation is not satisfiable:

$$\neg((\exists y \forall x\, p(x,y)) \Rightarrow (\forall x \exists y\, p(x,y))) \tag{7}$$
$$\neg(\neg(\exists y \forall x\, p(x,y)) \vee (\forall x \exists y\, p(x,y))) \tag{8}$$
$$(\exists y \forall x\, p(x,y)) \wedge \neg(\forall x \exists y\, p(x,y))) \tag{9}$$

So we have two clauses:

$$\exists y \forall x\, p(x,y) \tag{10}$$
$$\neg(\forall x \exists y\, p(x,y))) \tag{11}$$
$$\exists x \forall y \neg p(x,y))) \tag{12}$$

Skolemize (10) & (12):

$$\forall x\, p(x,C_1) \tag{13}$$
$$\forall y \neg p(C_2,y) \tag{14}$$

Dropping the universal quantifiers:

$$p(x,C_1) \tag{15}$$
$$\neg p(C_2,y) \tag{16}$$

Unify with $\theta = \{x/C_2, y/C_1\}$

Note that there was typo in the original question paper as follows:

$$(\forall x \exists y\, p(x,y)) \Rightarrow (\exists y \forall x\, p(x,y))$$

My apologies. Unfortunately, none of the students spotted this typo during the exam. Fortunately, the typo didn't really affect this question (which is perhaps why no one spotted it) because in the original formulation, one would be able to work out the problem similiarly. The only issue is that in the final step, the unification step is of the following form: $\theta = \{x/B(y), y/A(x)\}$, which is not permissible. Why not? :-)

It is however not obvious, especially under exam conditionst that this is wrong. Students who did this "wrong" thing got full credit.

Question 4 : Diagnosis of Lung Cancer [40 marks]

You have been employed by a hospital as a consultant for a new medical diagnosis project. The goal of the project is to try to diagnose lung cancer without invasive surgery. You are provided with the following data on 12 patients:

Patient	Smoker?	Greasy Diet?	Exercise?	Coughing?	Lung Cancer?
1	Yes	No	Yes	No	Yes
2	No	Yes	No	Yes	No
3	No	No	No	No	No
4	Yes	No	No	No	No
5	Yes	Yes	Yes	No	No
6	No	No	No	Yes	No
7	Yes	Yes	No	No	No
8	Yes	No	No	Yes	Yes
9	Yes	No	No	No	No
10	Yes	Yes	No	Yes	Yes
11	No	No	No	Yes	No
12	No	Yes	No	Yes	No

A. Construct a decision tree using information gain to predict whether a given patient is likely to have lung cancer. Is your decision tree *consistent*? Explain. [Note: You should justify your answer by showing your computations for information gain. More space for your answer is available on the next two pages.] [15 marks]

$$
\begin{aligned}
I(\frac{3}{12},\frac{9}{12}) &= -\frac{3}{12}\log_2\frac{3}{12}-\frac{9}{12}\log_2\frac{9}{12}\\
&= 0.811\\
Gain(Smoker) &= 0.811-[\frac{7}{12}I(\frac{3}{7},\frac{4}{7})+\frac{5}{12}I(1,0)]\\
&= 0.811-[\frac{7}{12}(-\frac{3}{7}\log_2\frac{3}{7}-\frac{4}{7}\log_2\frac{4}{7})]\\
&= 0.236\\
Gain(Greasy) &= 0.811-[\frac{5}{12}I(\frac{1}{5},\frac{4}{5})+\frac{7}{12}I(\frac{5}{7},\frac{2}{7})]\\
&= 0.811-[\frac{5}{12}(-\frac{1}{5}\log_2\frac{1}{5}-\frac{4}{5}\log_2\frac{4}{5})+\frac{7}{12}(-\frac{5}{7}\log_2\frac{5}{7}-\frac{2}{7}\log_2\frac{2}{7})]\\
&= 0.0067\\
Gain(Exercise) &= 0.811-[\frac{2}{12}I(\frac{1}{2},\frac{1}{2})+\frac{10}{12}I(\frac{2}{10},\frac{8}{10})]\\
&= 0.811-[\frac{2}{12}(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2})+\frac{10}{12}(-\frac{1}{5}\log_2\frac{1}{5}-\frac{4}{5}\log_2\frac{4}{5})]\\
&= 0.0430\\
Gain(Coughing) &= 0.811-[\frac{6}{12}I(\frac{2}{6},\frac{4}{6})+\frac{6}{12}I(\frac{1}{6},\frac{5}{6})]\\
&= 0.811-[\frac{6}{12}(-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3})+\frac{6}{12}(-\frac{1}{6}\log_2\frac{1}{6}-\frac{5}{6}\log_2\frac{5}{6})]\\
&= 0.0271
\end{aligned}
$$

Hence, the root node should be "Smoker?". Thereafter, we have the following sample points remaining for "Smoker? Yes":

Patient	Greasy Diet?	Exercise?	Coughing?	Lung Cancer?
1	No	Yes	No	Yes
4	No	No	No	No
5	Yes	Yes	No	No
7	Yes	No	No	No
8	No	No	Yes	Yes
9	No	No	No	No
10	Yes	No	Yes	Yes

$$
\begin{aligned}
I(\frac{3}{7},\frac{4}{7}) &= -\frac{3}{7}\log_2\frac{3}{7}-\frac{4}{7}\log_2\frac{4}{7}\\
&= 0.985\\
Gain(Greasy) &= 0.985-[\frac{3}{7}I(\frac{1}{3},\frac{2}{3})+\frac{4}{7}I(\frac{2}{4},\frac{2}{4})]
\end{aligned}
$$

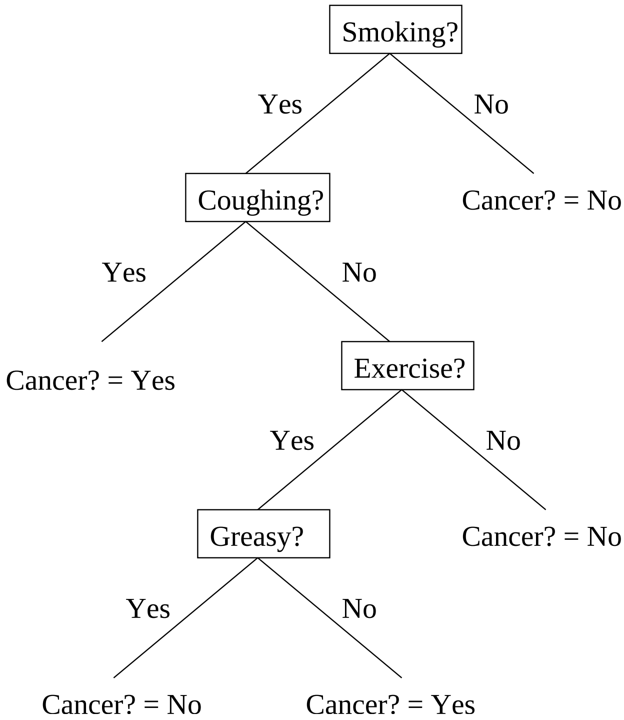
$$
\begin{aligned}
&= 0.985 - [\frac{3}{7}(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}) + \frac{4}{7}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}))] \\
&= 0.020 \\
Gain(Exercise) &= 0.985 - [\frac{2}{7}I(\frac{1}{2},\frac{1}{2}) + \frac{5}{7}I(\frac{2}{5},\frac{3}{5})] \\
&= 0.985 - [\frac{2}{7}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) + \frac{5}{7}(-\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}))] \\
&= 0.0059 \\
Gain(Coughing) &= 0.985 - [\frac{2}{7}I(1,0) + \frac{5}{7}I(\frac{1}{5},\frac{4}{5})] \\
&= 0.985 - [\frac{2}{7}(\frac{5}{7}(-\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5}))] \\
&= 0.470
\end{aligned}
$$

Hence, the next node should be "Coughing?". Thereafter, we have the following sample points remaining for "Smoker? No":

Patient	Greasy Diet?	Exercise?	Lung Cancer?
1	No	Yes	Yes
4	No	No	No
5	Yes	Yes	No
7	Yes	No	No
9	No	No	No

$$
\begin{aligned}
I(\frac{1}{5},\frac{4}{5}) &= -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} \\
&= 0.722 \\
Gain(Greasy) &= 0.722 - [\frac{2}{5}I(1,0) + \frac{3}{5}I(\frac{1}{3},\frac{2}{3})] \\
&= 0.722 - [\frac{3}{5}(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3})] \\
&= 0.171 \\
Gain(Exercise) &= 0.722 - [\frac{3}{5}I(1,0) + \frac{2}{5}I(\frac{1}{2},\frac{1}{2})] \\
&= 0.722 - [\frac{2}{5}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2})] \\
&= 0.322
\end{aligned}
$$

The following is the decision tree.



Yes, tree is consistent, i.e. fits all the data.

Note: This question is somewhat painful but truthfully, the student is not required to compute all the numbers above. Only the remainder matters and if the student understands this, he/she would only have to compute $4+3+2=9$ *numbers.*

B. From your decision tree, predict if the following patients have lung cancer: [2 marks]

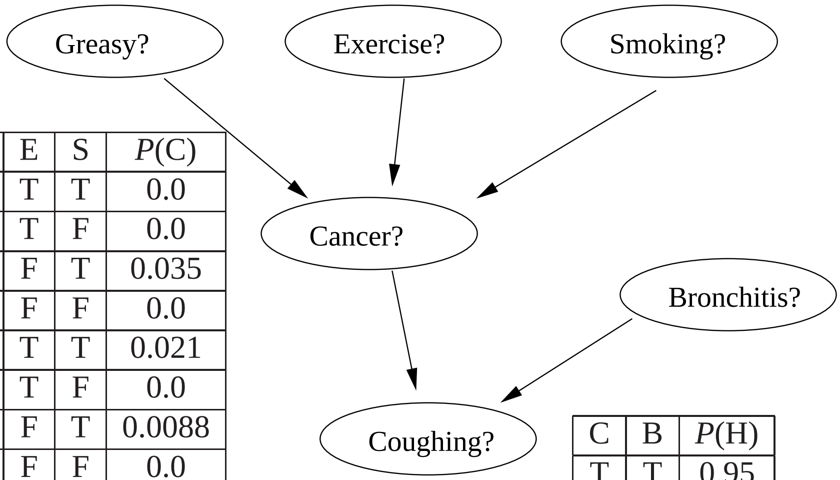
Patient	Smoker?	Greasy Diet?	Exercise?	Coughing?
13	Yes	No	No	Yes
14	Yes	Yes	Yes	No

Patient 13 is predicted to have lung cancer while patient 14 is not.

C. Suppose it eventually turns out that the diagnosis for the two patients in Part B above were different from those predicted by your decision tree. Suggest a possible reason why your decision tree may have come to the wrong conclusion. [3 marks]

It is possible that we are missing a critical attribute and that the last two attributes exercise and greasy do not actually affect the likelihood of cancer. Also, it is possible that there is a probabilitistic element that is not captured in the decision tree.

After some discussions with the doctors, you come to the conclusion that it might perhaps be a good idea to construct a probabilitistic model for diagnosis instead of using a decision tree. With some help, you construct the following Bayesian network:



P(G)
0.2

P(E)
0.3

P(S)
0.2

G	E	S	P(C)
T	T	T	0.0
T	T	F	0.0
T	F	T	0.035
T	F	F	0.0
F	T	T	0.021
F	T	F	0.0
F	F	T	0.0088
F	F	F	0.0

P(B)
0.01

C	B	P(H)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.01

We represent the variables Greasy?, Exercise?, Smoking?, Cancer?, Cough? and Bronchitis? with G, E, S, C, H and B respectively.

D. As shown in the figure, G, E, S, and B are independent variables and their prior probabilities are known. Also, you managed to find the conditional probabilities for H in a medical journal. The only piece of data missing are the conditional probabilities for C. Suppose you are told that in addition to the 12 patients listed in Part A, there are another 988 persons who are not patients (i.e. do not have cancer), and together, these 1,000 persons form a good sample of the population. Derive a reasonable estimate for the conditional probability table for C. You can explain your methodology and working in the space below. [6 marks]

We use the priors for G, E and S to estimate the populations of the healthy persons. Then we add the patient count from Part A.

The point of this question is to have students understand that we sometimes have to estimate the conditional probability table from simple statistics and sometimes have to make assumptions.

G	*E*	*S*	*Healthy Count*	*Patient Count*	*Cancer Count*	*Probability*
T	*T*	*T*	*11.86*	*1*	*0*	*0.0*
T	*T*	*F*	*47.42*	*0*	*0*	*0.0*
T	*F*	*T*	*27.66*	*2*	*1*	*0.035*
T	*F*	*F*	*110.66*	*2*	*0*	*0.0*
F	*T*	*T*	*47.42*	*1*	*1*	*0.021*
F	*T*	*F*	*189.70*	*0*	*0*	*0.0*
F	*F*	*T*	*110.66*	*3*	*1*	*0.0088*
F	*F*	*F*	*442.62*	*3*	*0*	*0.0*

E. Suppose a patient is coughing, what is the probability that he has lung cancer? [6 marks]
This part and the next part test the student's ability to come up with the correct expressions from the available conditional probability tables. Even if the estimated conditional probabilities from the previous part are wrong, full credit will be given if the correct expressions are used.

$$
\begin{aligned}
P(H|C) &= P(H|C,B)P(B) + P(H|C,\neg B)P(\neg B) \\
&= 0.95 \times 0.01 + 0.8 \times 0.99 \\
&= 0.8015 \\
P(C) &= P(C|G,\neg E,\neg S)P(G)P(\neg E)P(\neg S) + P(C|\neg G,E,S)P(\neg G)P(E)P(S) \\
&\quad + P(C|\neg G,\neg E,S)P(\neg G)P(\neg E)P(S) \\
&= 0.035 \times 0.2 \times 0.7 \times 0.8 + 0.021 \times 0.8 \times 0.3 \times 0.2 \\
&\quad + 0.0088 \times 0.8 \times 0.7 \times 0.2 \\
&= 0.0059 \\
P(H) &= P(H|C,B)P(C)P(B) + P(H|\neg C,B)P(\neg C)P(B) \\
&\quad + P(H|C,\neg B)P(C)P(\neg B) + P(H|\neg C,\neg B)P(\neg C)P(\neg B) \\
&= 0.95 \times 0.0059 \times 0.01 + 0.9 \times 0.995 \times 0.01 \\
&\quad + 0.8 \times 0.0059 \times 0.99 + 0.01 \times 0.995 \times 0.99 \\
&= 0.0235 \\
P(C|H) &= \frac{P(H|C)P(C)}{P(H)} \\
&= \frac{0.8015 \times 0.0059}{0.0235} \\
&= 0.2
\end{aligned}
$$

One student estimated $P(C)$ as $\frac{3}{1000}$ and proceeded to solve this question. Full credit was given.

F. Suppose in addition to the cough, you are told that this patient is also suffering from bronchitis, what is the probability that he has lung cancer now? [6 marks]

$$
\begin{aligned}
P(H|B) &= P(H|B,C)P(C) + P(H|B,\neg C)P(\neg C) \\
&= 0.95 \times 0.0059 + 0.9 \times 0.995 \\
&= 0.9 \\
P(C|H,B) &= \frac{P(C,H,B)}{P(H,B)} \\
&= \frac{P(H|B,C)P(B)P(C)}{P(B)P(H|B)} \\
&= \frac{P(H|B,C)P(C)}{P(H|B)} \\
&= \frac{0.95 \times 0.0059}{0.9} \\
&= 0.0062
\end{aligned}
$$

G. Compare the answers in Parts E and F. What is the intuitive reason for the difference? [2 marks]

It makes sense that the answer in Part F is less than that in Part E. Basically, because a cough is one symptom of lung cancer, we would expect a person with a cough to have some probability of suffering from cancer. However, bronchitis can also result in a cough, so when told that a person has bronchitis as well, the bronchitis "explains" for the cough and hence it makes it less likely that the cough is the symptom of cancer and hence the probability of cancer goes down.

—— HAPPY HOLIDAYS! ——