DSO 570: Solution to Homework 2

Learning Objectives Assessed:

- Identify types of uncertainty that can be appropriately modeled as random variables. (Model)
- Apply basic concepts in probability to solve problems. (Analyze)
- Apply a decision tree to model sequential decision making and find the optimal solution. (Model and Analyze)

1. For each of the following uncertainties, decide whether or not you would model it as a random variable. If so, describe briefly how will you go about estimating the underlying probabilities. If not, describe why it would not be appropriate and what might be a better alternative to cope with the uncertainty? (There may not be a definitive "right answer." This question assesses how you think. What's important is the arguments you make.)
The key consideration in whether it makes sense to model something as a random variable is whether we have data or understanding to obtain reasonable estimate of the underlying probabilities. When we don't understand an uncertainty at all and don't have any data to estimate what may happen, it doesn't make sense to make up numbers and model the uncertainty as a random variable. For those kind of uncertainty, it is better to consider a "robust optimization" approach, which is to make sure that the decision we choose is reasonable under all possible contingencies.

a. The Dow Jone Industrial Average exactly one month from today.
There is much data on the Down Jone Industrial Average. Although the value exactly one month from today is unknown, we can estimate a distribution based on historical data, especially as we incorporate current trends and macro economic indicators that are correlated with the stock.

b. Whether or not there will be a nuclear holocaust in the next five years.
There is no prior nuclear holocaust in the past, and no clear understanding of what would cause such an event. Moreover, even if one obtains a probabilistic estimate, the probability would be very small. However, the consequence of such an event is so large, that despite the small likelihood of happening, we would like to prevent it. Hence, I would argue that one should not model this uncertainty as a random variable, but as an unlikely but scary outcome that one would do all that is possible to prevent.

c. The outcome of the next presidential election.
There is much historical data on presidential elections in the US, as well as more recent poll data. Hence, one can form reasonable estimates of the likelihood of each outcome, so this is a type of uncertainty that can be appropriately modelled as a random variable.

d. (Suppose you are part of the airport security defending against terrorist attacks, and whatever you decide might be leaked to the terrorists.) Where in the airport will the next attack will take place.
In this case, although there is some data on prior terrorist attacks, and one may form conjectures of most vulnerable places for an attack, it is not appropriate to model as a random variable for the following reason: whatever we believe about the behavior of terrorists may be leaked to the terrorists, and they might adapt their behavior and attack where we are not guarding. Therefore, because our probabilistic modeling has the potential to affect the actual event, it is better to find defense strategies that are robust to any possible attack, rather than to try to predict likely attacks and prioritize them (at risk of leaving other vulnerabilities).

2. (DMD Exercise 2.1) A four-sided die is engraved with the numbers 1 through 4 on its four different sides. Suppose that when rolled, each side (and hence each number) has an equal probability of being the bottom face when it lands. We roll two such dice. Let X be the sum of the numbers on the bottom faces of the two dice.

a. What is the probability that X is at least five?

Each four-sided die is equally likely to be 1 through 4, and the two dice are independent of each other. Therefore the set of possible outcomes are the following, each with probability 1/16.

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

$P(X \geq 5)$ is the event of having the sum of dice at least 5; where X denotes the sum of the sides. Of the above outcomes, there are 10 with sum greater than or equal to 5.

$P(X \geq 5)$ = P(1,4) + P(2,3) + P(2,4) + P(3,2) + P(3,3) + P(3,4) + P(4,1) + P(4,2) + P(4,3) + P(4,4) = **10/16**.

b. How does your answer to (a) change if you are told that the bottom face of the first die has the number 3 on it?

Probability of the sum being at least 5 given the first dice is a 3 can be represented as

$$P(X \geq 5\,|first\ is\ 3) = \frac{P(X\geq 5\ and\ first\ is\ 3)}{P(first\ is\ 3)} = \frac{3/16}{4/16} = \frac{3}{4}$$

c. How does your answer to (a) change if you are told that the bottom face of one of the dice has the number 3 on it?

Probability of the sum being at least 5 given one of the dice has the number 3

$$P(X \geq 5\,|\ at\ least\ 3\ in\ one\ die) = \frac{P(X\geq 5\ and\ at\ least\ 3\ in\ one\ die)}{P(At\ least\ 3\ in\ one\ die)} = \frac{5/16}{7/16} = \frac{5}{7}$$

3. (DMD Exercise 2.2) We toss a coin three times. Let the outcome of this experiment be the sequence of heads (H) and tails (T) resulting from the three tosses.

a. Enumerate all of the possible outcomes of this experiment.

Tossing a coin has two outcomes {H, T}. Tossing a coin three times has 2^3 possibilities i.e. $\{(HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), (TTT)\}$

b. What is the probability of the outcome "HHT"?

For this question we could simply count the number of times we have "HHT" in our possibilities

$P(HHT) = 1/8. \qquad \{(HHT)\}$

c. What is the probability of the event "The first two tosses resulted in heads"?

Two out of eight possibilities have first two results as heads
$P(First\ two\ tosses\ are\ heads)$ = 2/8. $\{(HHH), (HHT)\}$

d. What is the probability of the event "There were two heads in a row among the three tosses"?

Three out of eight possibilities have heads in a row

$P(Two\ heads\ in\ a\ row)$ = 3/8. $\{(HHH), (HHT), (THH)\}$

4. (DMD Exercise 2.4) An oil company is drilling for oil at three promising sites. According to geological tests, the probabilities of finding oil at these three sites are 0.70, 0.85, and 0.80, respectively. The presence of oil at any one of the sites is presumed to be independent of the presence of oil at any of the other sites.

Let the three sites be A, B, C.

Probability of finding oil at site A is given as $P(A) = 0.70$
Probability of finding oil at site B is given as $P(B) = 0.85$
Probability of finding oil at site C is given as $P(C) = 0.80$

a. What is the probability of finding oil at all three of the sites?

Because the events A, B and C are all independent, the probability of finding oil at all three sites is
$P(A\ and\ B\ and\ C) = P(A) * P(B) * P(C) = 0.70 * 0.85 * 0.80 = 0.476$

b. What is the probability of not finding oil at any of the three sites?

Let A' denote the event A does not happen. Similarly for B' and C'. The probability of A not happening is simply 1-P(A). Hence, we can multiple the "not happening" probabilities together as before:

$P(A'\ and\ B'\ and\ C') = (1 - P(A)) * (1 - P(B)) * (1 - P(C)) = 0.30 * 0.15 * 0.20 = 0.009$

5. (DMD Exercise 2.8) On a television game show, there are three boxes. Inside one of the boxes there is a check for 10,000 dollars. If you pick the box that contains the check, you keep the money. Suppose you pick one of the boxes at random. The host of the game opens one of the other boxes and reveals that it is empty. The host then offers you the chance to change your pick. Should you change your pick? If so, why? If not, why not?

Label the first box we pick as box A, and the other two as boxes B and C. At this point, the situation is symmetric, hence P(A)=P(B)=P(C)=1/3.

Define two additional event: "Staying wins" or "Switching wins," corresponding to whether or not in retrospect we should have stayed with A or have switched. The joint probability table is as follows.

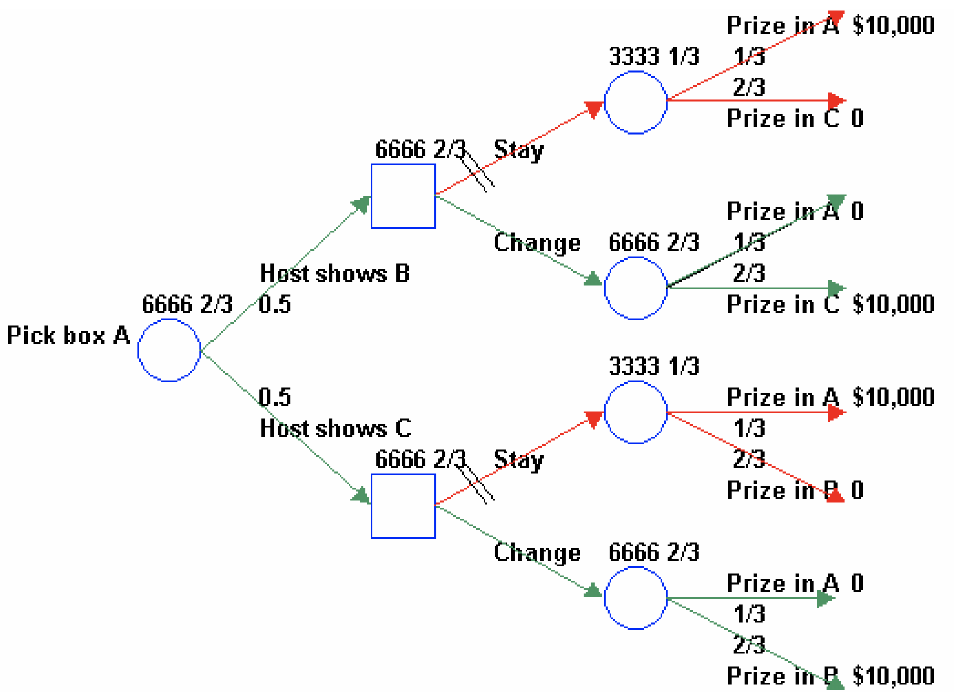
	A	B	C
Staying wins	1/3	0	0
Switching wins	0	1/3	1/3

In other words, whenever event A happens, staying wins. However, whenever event B or C happens, switching wins. This is because if the true box were B or C, then switching always wins, because **the host's opening of an empty box among B and C helps you to choose the right box among B and C**.

The total probability of "staying wins" is 1/3 and the probability of "switching wins" is 2/3, so we should switch.

[Alternate Solution]

Consider the decision tree below. We assume that the boxes are labeled A, B, and C. Without loss of generality, we assume that the initially chosen box is A. Hence, there are four possible scenarios, each with the same probability: the prize is in A and the host shows B, the prize is in A and the host shows C, the prize is in B and the host must show C, or the prize is in C and the host must show B. Therefore, the host shows B with probability 2/4 = 0.50, and it shows C with probability 2/4 = 0.50. According to this decision tree, you should always change your pick.



6. (DMD Exercise 2.10) A hardware store has received two shipments of halogen lamps. The first shipment contains 100 lamps, 4% of which are defective. The second shipment contains 50 lamps, 6% of which are defective. Suppose that Emanuel picks a lamp (at random) off of the shelf and purchases it, and he later discovers that the lamp he purchased is defective. Is the defective lamp more likely to come from the first shipment or from the second shipment?

Let S1 denote the lamp comes from 1st shipment,
Let S2 denote the lamp comes from 2nd shipment,
Let D denote the lamp is defective.

To determine 'whether the defective lamp is more likely to come from the first shipment or from the second shipment', we want to compare $P(S1 \mid D)$ to $P(S2 \mid D)$.

Given: $P(S1) = \frac{100}{150} = \frac{2}{3}$; $P(S2) = \frac{50}{150} = \frac{1}{3}$; $P(D \mid S1) = 0.04$, ; $P(D \mid S2) = 0.06$.

Probability of lamb being defective can be found by:
P(D) $= P(D|S1) * P(S1) + P(D|S2) * P(S2) = 0.04 * \frac{2}{3} + 0.06 * \frac{1}{3} = \frac{7}{150}$.

Therefore:

$P(S1 \mid D) = \frac{P(D \mid S1) * P(S1)}{P(D)} = \frac{8}{14}$ and $P(S2 \mid D) = \frac{P(D \mid S2) * P(S2)}{P(D)} = \frac{6}{14}$.

In conclusion, the lamp is more likely to come from the first shipment.

7. (DMD Exercise 2.11) It is estimated that one third of the population in a given county is infected with the tuberculosis (TB) bacteria. The human body is usually able to successfully fight the TB bacteria and so prevent the onset of the TB disease. Consequently, a person infected with the TB bacteria has only a 10% chance of development the TB disease over his/her lifetime.

Let TB denote the event that the person is infected with the TB bacteria.
Let D denote the event that the person has the disease.

Let TB' and D' denote the complementary event from the above (complementary means opposite outcome.)

Given: $P(TB) = \frac{1}{3}$ and $P(D \mid TB) = 0.10$.

a. Suppose that we choose a person at random from all of the people in the county. What is the probability that this person has the TB disease?

$$P(D) = P(D|TB) * P(TB) + P(D|TB') * P(TB')$$

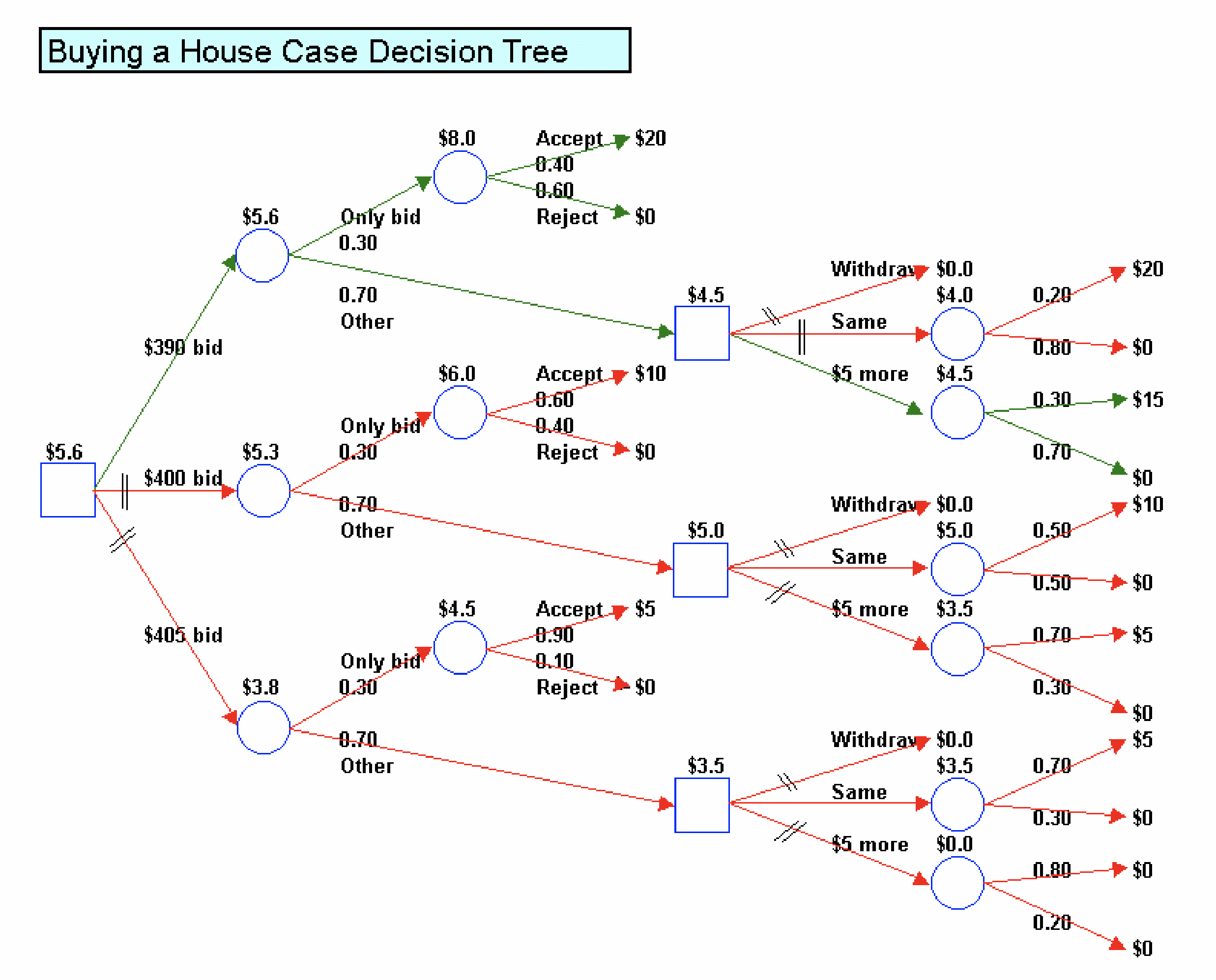
$$P(D) = 0.10 * \left(\frac{1}{3}\right) + 0 * \left(\frac{2}{3}\right) = \frac{1}{30}$$

b. Suppose that among those people who have died in the county in the last year, we perform an autopsy on one of the bodies chosen at random, and we find that this person did not have the TB disease. What is the probability that this person had been infected with the TB bacteria?

$$P(TB \mid D') = \frac{P(D'|TB) * P(TB)}{P(D')} = \frac{9}{29}$$

8. (DMD Section 1.6, "Buying a House" case) This problem begins with the text "Debbie and George Calvert are thinking of making an offer to purchase a house..."

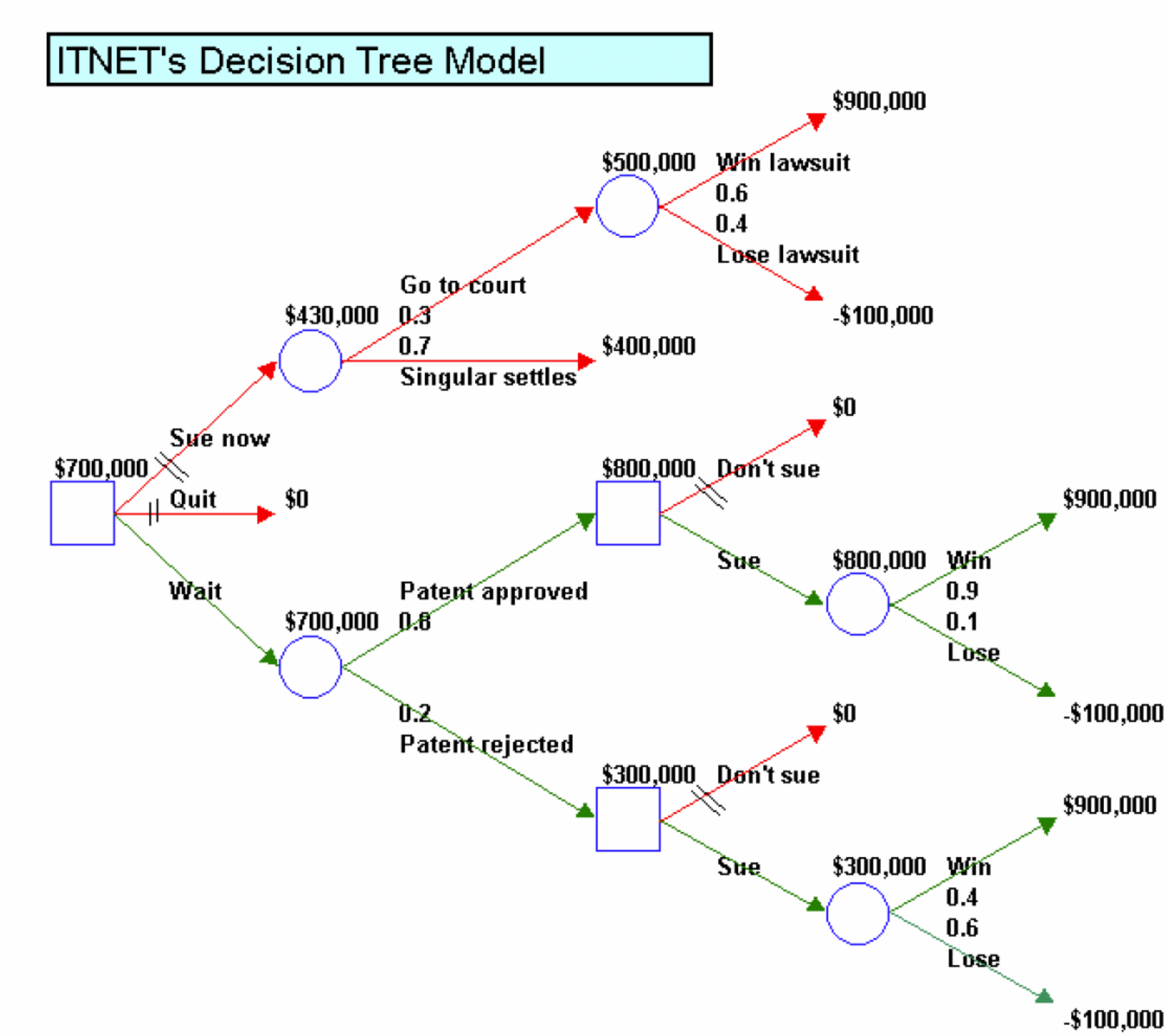
a. Decision Tree



b. The optimal decision strategy for Debbie and George is to bid $390,000 initially. If there are other bidders and the seller ask them to submit a final offer the following day, then they should increase their offer by $5,000 the next day, that is, make a final offer of $395,000. The EMV for this strategy is $5,600.

9. (DMD Exercise 1.4) This problem begins with the text "Anders and Michael were classmates in college. In their spare time while undergraduates, they developed a software product that regulates traffic on Internet sites..."

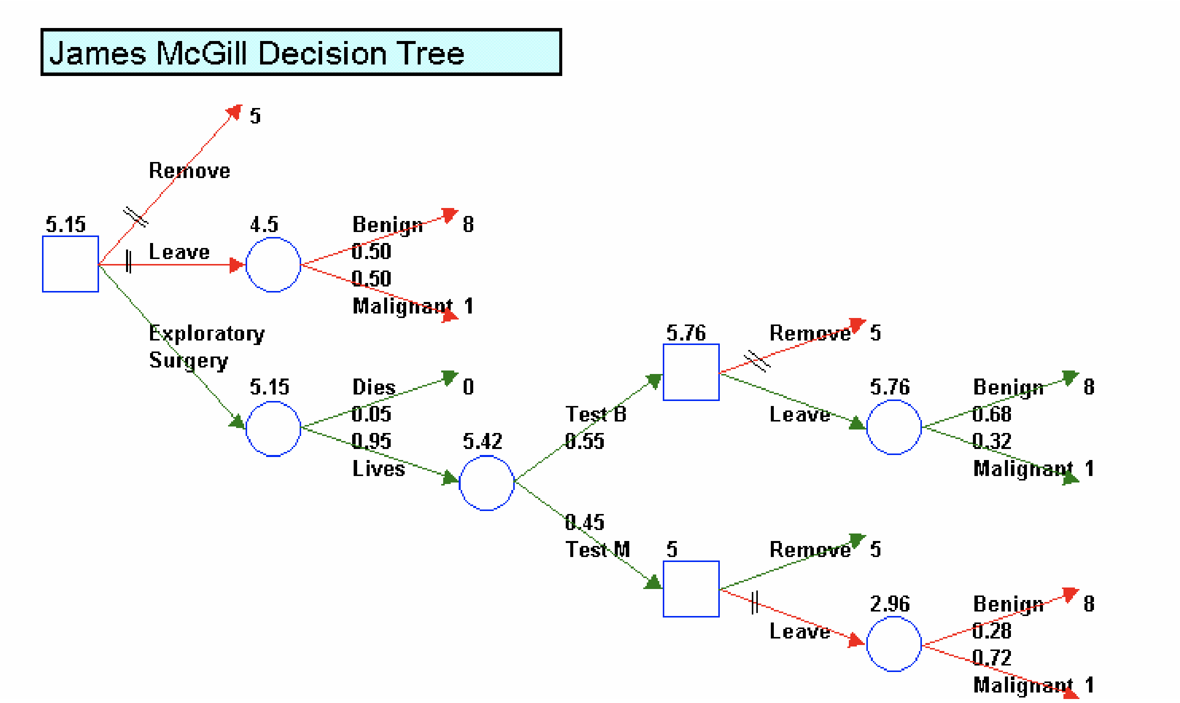
a. Decision Tree



b. ITNET's optimal decision strategy is to wait for the result of their patent application, and then, sue Singular regardless of whether the patent is approved or not.

10. (DMD Exercise 1.6) This problem begins with the text "James McGill, age 68, was recently diagnosed with a particular type of brain tumor..."

a. Decision Tree



b. The missing probabilities are for:

- The event that the tumor is benign given that the exploratory surgery indicated a benign tumor.
- The event that tumor is benign given that the exploratory surgery indicated a malignant tumor.
- The exploratory surgery will indicate a benign tumor.

c. The corresponding probabilities are shown in the table below. James' optimal decision strategy is to undergo exploratory surgery. If he survives and the surgery indicates a benign tumor, he should choose to leave the tumor. If he survives and the surgery indicates a malignant tumor, he should choose to remove the tumor. The expected life years from this strategy is 5.15 years.

Joint Probabilities

	Benign	Malignant	
Test Benign	0.375	0.175	0.55
Test Malignant	0.125	0.325	0.45
	0.50	0.50	

Conditional Probabilities:
P(Benign| Test Benign) = 0.375/0.55 = 0.682
P(Benign| Test Malignant) = 0.125/0.45 = 0.278

d. In this case we value the outcomes not by remaining number of life years, but by whether that expected numbers is greater than or equal to 2. In other words, if the remaining life years is less than 2, we assign a value of 0. If the remaining life years is at least 2, we assign a value of 1. Hence, we are no longer maximizing the expected number of life years, but maximizing the expected value of the indicator variable we defined, which is the same as the probability that he lives at least 2 years. By doing this, we find that James' optimal decision strategy is to remove the tumor without undergoing exploratory surgery, as this would (according to the assumptions of the case) guarantee 5 years of life.

e. This problem poses several ethical questions:
   - How do we value a medical treatment for a patient? Is it by how many years it gives to a patient? Or is it by quality of life? If the former, then we are saying that it is okay to have patients suffer as long as they live longer (although cancer treatment might result in a miserable existence). If the latter, then how do we define quality? (As can be seen in part d, what a patient cares about may be very subjective.)
   - What about the tradeoff between patient outcome and societal costs? Surgery and other operations can be very expensive, with the brunt of the burden paid by the health insurance system from the pocket of healthier patients. How much are we as a society willing to pay for an additional year of life?
   - When we advise patients based on analytics methods such as decision trees, do they truly understand the implications? For example, if we had advised James McGill based on the first analysis on the expected years of life remaining, this may have persuaded him to make one choice, but he may not understand the implications. One must keep in mind that the expected value is a number that should be interpreted in a long run sense. For instance, suppose that you buy a $1 lottery ticket with a probability of 1:10000000 of winning 100 million dollars, then your expected win is $9. This means that if you play this game many times, in the long run your average number of earnings is $9. Nevertheless, it is very likely that you will play for a lifetime and never win this lottery. In this case, exploratory surgery gives him a higher expected number of years remaining, at the risk of a higher chance of dying within 2 years and not seeing his grandchildren.