15.063 Communicating with Data
Summer 2003

Homework Assignment #1

Issued: Lecture 3. **Due: Lecture 7, before Lecture.**

1. Decision Trees

Extensive logging has exposed a hillside in San Carlos to the possibility of a mudslide. Reforestation is underway, but it will be a year before the new vegetation will be mature enough to remove the danger. If a slide occurs in the interim, human injuries will be avoided because mud moves slowly. The damage from such a slide would be limited to the road that passes beneath the hill. Construction of a retaining wall on the uphill side of the road has been suggested as a possible step to prevent this damage.

The Mayor of San Carlos is puzzled by the uncertainty concerning the issue. He has consulted with an expert who states that there is only one chance in 100 that a slide will occur within the next year. The expert adds that roughly 5% of all such slides break through retaining walls like the one proposed. The retaining wall would cost $40,000 to build. The road would cost about $1,000,000 to repair if damaged by a mudslide. The expert points out that she can better assess the likelihood of a slide occurring in the next year if she conducts a geological test of the igneous rock layer below the hillside. Like any test, this one is imperfect. A positive test outcome would indicate a higher chance of a slide than a negative test outcome. The test has been conducted at sites at which slides eventually occurred and at sites at which slides did not subsequently occur. The information from these previous tests can be summarized as follows. Positive test outcomes had been reported on 90% of the sites at which slides subsequently occurred. Negative test results had been reported at 85% of the sites at which slides did not subsequently occur.

Assignment: As an aide to the mayor of San Carlos, what action would you recommend be taken? Please explain your results.
Note 1: assume that the cost of the test is minimal.
Note 2: you may use *Treeplan* or construct the decision tree yourself.

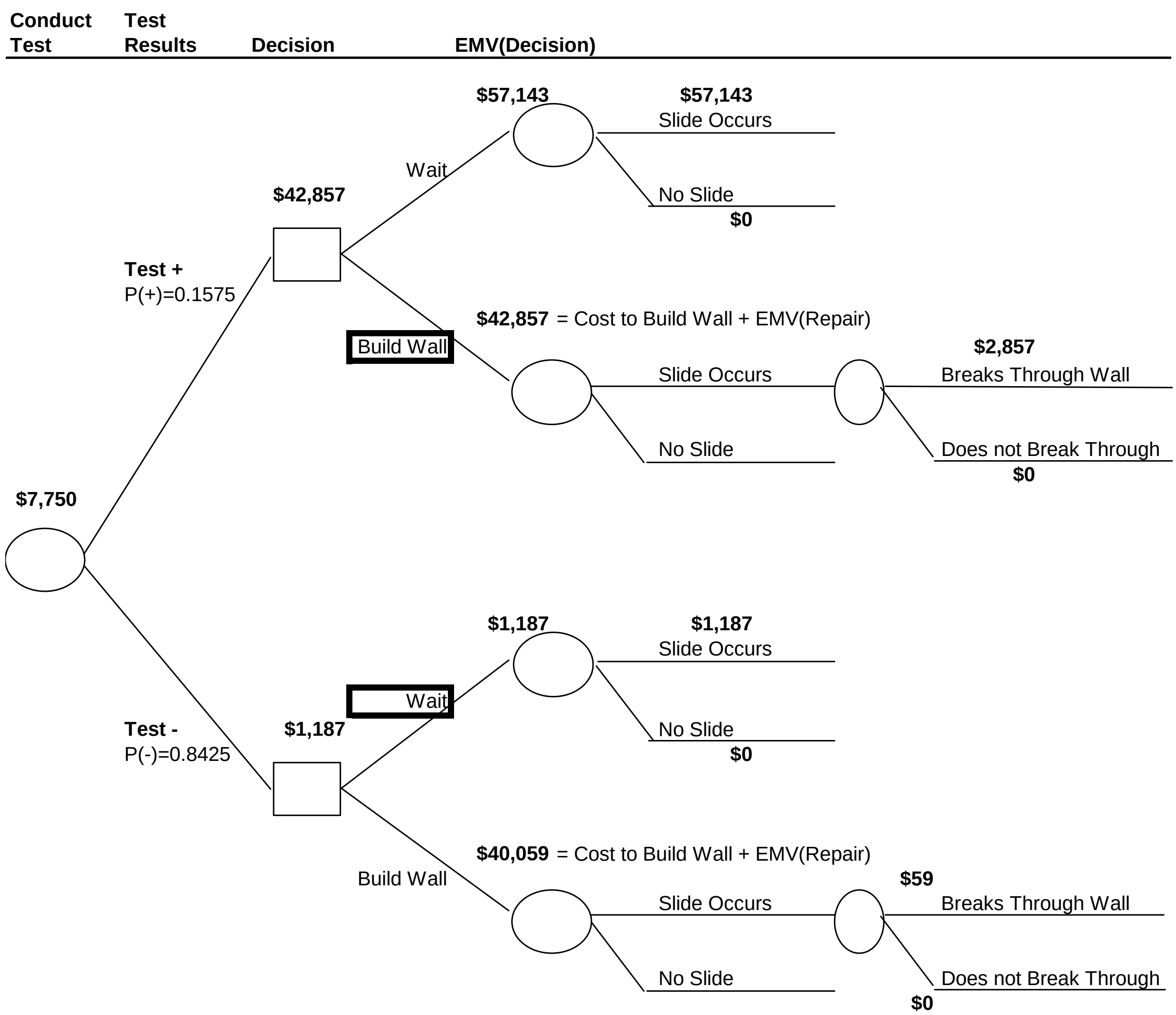
***Solution*:** From the statement of the problem we know the following.

- p(Slide) = 0.01
- p(Test + | Slide) = 0.90
- p(Test - | no Slide) = 0.85
- p(break through wall | Slide) = 0.05

From there, we compute the probability table:

	Test +	Test -	Total
Slide	0.8415	0.1485	0.99
no Slide	0.001	0.009	0.01
Total	0.1575	0.8425	1

The conditional probabilities can be computed from the table directly. Solving the decision tree gives that they should build the wall if the test is positive and not build it otherwise. The expected cost they will face is $7,750.



2. Laws of Probability

Exercise 2.3 from the book, *Data, Models, and Decisions*: *The Fundamentals of Management Science* by Dimitris Bertsimas and Robert M. Freund, Southwestern College Publishing, 2000.

Please explain your results.

Solution:

(a) p(at least 4 years experience) = p(4 years experience) + p(5 or more years experience)
= 15/100 + 35/100 = 50/100.

(b) p(at least 4 years experience | at least 3 years experience)
= p(at least 4 years experience) / p(at least 3 years experience) =
= (50/100) / (80/100) = 50/80.

3. Conditional Probability

To the best of our knowledge, with probability 0.8 Al is guilty of the crime for which he is about to be tried. Bo and Ci, each of whom knows whether or not Al is guilty, have been called to testify. Bo is a friend of Al's and will tell the truth if Al is innocent, but will lie with probability 0.2 if Al is guilty. Ci hates everybody but the judge and will tell the truth if Al is guilty but will lie with probability 0.3 if Al is innocent. Given this model:

a. Determine the probability that the witnesses give conflicting testimony.
b. Which witness is more likely to commit perjury?
c. What is the conditional probability that Al is innocent, given that Bo and Ci gave conflicting testimony?
d. Are the events "Bo tells a lie" and "Ci tells a lie" independent? Are these events conditionally independent to an observer who knows whether or not Al is guilty?

Solution:
We will provide two solutions to this problem because we realize that there are different ways to solve a problem. In particular, as this was a hard one, it may be instructive to understand both. The answer is long because I wrote this with all the details, but hopefully you will be able to follow all the reasoning.

Solution A:

a. Let A be the event "Al is guilty" and X be the event "there is conflicting testimony". The motivation for defining event X is that what we go after is its probability, i.e., p(X). Writing what we know mathematically gives:

- p(A) = 0.8
- p(not A) = 0.2

Now, let's try to complete a probability table with columns A and not A and rows X and not X.

In order to do that, we determine the conditional probabilities.

Let's start by conditioning on A (i.e., we restrict ourselves to the case when Al was guilty). In that case, we know that Al was guilty. Therefore, Bo can potentially lie but Ci will always say the truth. The only way to have conflicting testimony is when Bo says that Al was innocent (Bo lies) and that can happen with probability 0.2. With symbols, we just deduced that p(X|A)=0.2

Equivalently, we condition now on "not A" (i.e., we restrict ourselves to the case when Al was innocent). In that case, we know that Al was innocent. Therefore, Ci can potentially lie but Bo will always say the truth. The only way to have conflicting testimony is when Ci says that Al was guilty (Ci lies) and that can happen with probability 0.3. With symbols, we just deduced that p(X|not A)=0.3.

With these probabilities, we are ready to fill in the table.

	A	not A	Total
X	0.2 x 0.8 = 0.16	0.3 x 0.2 = 0.06	0.22
not X	0.8 – 0.16 = 0.64	0.2 – 0.06 = 0.14	0.78
Total	0.8 (from bullets)	0.2 (from bullets)	1

Last, we can read the answer directly from the table: 0.22.

b. To determine the probability Bo lies, we condition on A and not A (Al is guilty or not) and make table as in the previous part.

p(Bo lies|A)=0.2 and p(Bo lies|not A)=0 (Bo can only lie if Al is guilty).

	A	not A	Total
Bo lies	0.2 x 0.8 = 0.16	0 x 0.2 = 0	0.16
Bo does not lie	0.8 – 0.16 = 0.64	0.2 – 0 = 0.2	0.84
Total	0.8 (from bullets)	0.2 (from bullets)	1

For Ci, we do the same:
p(Ci lies|A)=0 and p(Ci lies|not A)=0.3 (Ci can only lie if Al is innocent).

	A	not A	Total
Ci lies	0 x 0.8 = 0	0.3 x 0.2 = 0.06	0.06
Ci does not lie	0.8 – 0 = 0.8	0.2 – 0.06 = 0.14	0.94
Total	0.8 (from bullets)	0.2 (from bullets)	1

From the two tables we see that is it more likely that Bo lies.

c. We are asked compute the conditional probability that Al is innocent, given that Bo and Ci gave conflicting testimony, i.e., p(not A|X). Using the third law of probability and reading those probabilities from the table that we constructed in part **a**, this is equal to
p(not A and X) / p(X) = 0.06 / 0.22 = 3/11

d. This was the most difficult part because it defies intuition. Let's see if "Bo tells a lie" is independent of "Ci tells a lie". The only way to do that is to verify if the equation

$$p(\text{Bo tells a lie and Ci tells a lie}) = p(\text{Bo tells a lie})\, p(\text{Ci tells a lie})$$

is true or not. If true, then the events are independent. If false, they are not. Let's verify that.
We know the RHS already, so we only have to compute the LHS: we proceed as in part **a**., computing the conditional probabilities.

Let's start by conditioning on A (i.e., we restrict ourselves to the case when Al was guilty). In that case, p(Bo tells a lie and Ci tells a lie | A) = 0 because as Al was guilty, Ci always says the truth.

Equivalently, we condition now on "not A" (i.e., we restrict ourselves to the case when Al was innocent). In that case, p(Bo tells a lie and Ci tells a lie | not A) = 0 because as Al was innocent, Bo always says the truth.

Therefore, we get that

$0 = p(\text{Bo tells a lie and Ci tells a lie}) \neq p(\text{Bo tells a lie})\, p(\text{Ci tells a lie}) = 0.16 \times 0.06$
from where we see that the two event are **not** independent.

For the second question in this part, we must prove independence conditioned on the knowledge that Al is guilty or not. We'll solve it for "Al guilty", the other is similar.

Let's see if "Bo tells a lie" is independent of "Ci tells a lie" when we condition on "Al is guilty". The only way to do that is to verify if the equation

$$p(\text{ Bo tells a lie and Ci tells a lie} \mid \text{Al is guilty }) = \\ p(\text{ Bo tells a lie } \mid \text{Al is guilty }) \times p(\text{ Ci tells a lie} \mid \text{Al is guilty})$$

is true or not. If true, then the events are independent conditioned on "Al is guilty". If false, they are not. Let's verify that.

The LHS was computed to be 0 before (in this question). Using the tables in part **b**., we can see that

$$p(\text{ Bo tells a lie } \mid \text{Al is guilty }) = p(\text{ Bo tells a lie } \mid A) = 0.16 / 0.8$$
$$p(\text{ Ci tells a lie} \mid \text{Al is guilty}) = p(\text{ Ci tells a lie} \mid A) = 0$$

Putting all that together, both sides of the equation we wrote down a few lines above are 0 and therefore the events are conditionally independent.

Solution B:

These are alternative solutions for parts **a**. and **c**. Let A be the event “Al is guilty”, B be “Bo says Al is guilty” and C be “Ci says Al is guilty”. Writing what we know mathematically gives:

- p(A) = 0.8
- p(B / not A) = 0 (if Al is not guilty, Bo will say the truth)
- p(B / A) = 0.8 (if Al is guilty, Bo will say so with probability 0.8)
- p(C / A) = 1 (if Al is guilty, Ci will say the truth)
- p(C / not A) = 0.3 (if Al is not guilty, Ci will say he is with probability 0.3)

With that we can now compute the probability tables:

	A	not A	Total
B	0.64	0.00	0.64
not B	0.16	0.20	0.36
Total	0.80	0.20	1.00

	A	not A	Total
C	0.80	0.06	0.86
not C	0.00	0.14	0.14
Total	0.80	0.20	1.00

a. The event that represents conflicting testimony is “(B and not C) or (not B and C)”. First notice that “B and not C” can never happen because if Ci said that Al was innocent, we know that he has to be innocent (but the judge doesn’t) and Bo will have said that Al is innocent too! Therefore, we need to compute p(not B and C). As B and C are *not* independent we cannot multiple them. Instead, we condition on the outcome of A:

p(not B and C) = p(not B and C and A) + p(not B and C and not A)
= p(not B and C | A) p(A) + p(not B and C | not A) p(not A).

After conditioning, B and C *are* independent (see part **d**. of Solution A), so rewriting we get:

p(not B and C) = p(not B | A) p(C | A) p(A) + p(not B | not A) p(C | not A) p(not A) =
p(not B and A) p(C and A) / p(A) + p(not B and not A) p(C and not A) / p(not A) =
0.16 x 0.80 / 0.80 + 0.20 x 0.06 / 0.20 = 0.22.

c. Reasoning as we did for the first part,

p(not A | not B and C) = p(not A and not B and C) / p(not B and C)
= p(not B and C | not A) p(not A) / p(not B and C).

As B and C are independent when conditioned on A, the last is equal to

p(not A | not B and C) = p(not B | not A) p(C | not A) p(not A) / p(not B and C)
= p(not B and not A) p(C and not A) / p(not A) / p(not B and C)
= 0.20 x 0.06 / 0.20 / 0.22 = 3/11

4. Gambling

A gambler has in his pocket a fair coin and a two-headed coin.

a. He selects one of the coins at random, and when he flips it, it shows heads. What is the probability that it is the fair coin?

b. Suppose that he flips the same coin a second time and again it shows heads. What is now the probability that it is the fair coin?

c. Suppose that he flips the same coin a third time and it shows tails. What is now the probability that it is the fair coin?

Solution:

Let the event **H** be "the coin showed heads", the event **F** be the "fair coin was selected" and the event **U** be "the two-headed coin was selected".

a. p(F|H) = p(F and H) / p(H) = p(H|F) P(F) / (p(H|F) P(F) + p(H|U) P(U)) =
0.5 x 0.5 / (0.5 x 0.5 + 0.5) = 1/3. (Alternatively, you can draw a probability table.)

b. Let the event **HH** be "the coin showed heads twice". Then:
p(F|HH) = p(F and HH) / p(HH) = p(HH|F) P(F) / (p(HH|F) P(F) + p(HH|U) P(U)) =
0.25 x 0.5 / (0.25 x 0.5 + 0.5) = 1/5.

c. Then it cannot have been the coin with two heads, so the probability is 1.0.