Exercises 8 · Probability and decision trees

Due Monday, April 20, 2014

Problem 1: conditional probability

You have just been appointed a Senior Vice President in the Preventive Care division of BlueCross/BlueShield of Texas, one of the state's largest health-insurance companies. Your first assignment is to study the cost-effectiveness of instituting a universal test for SOS, which stands for Seasonal-Onset Sooneritis. Your firm is thinking about making the test free and mandatory for all 10 million of its clients in Texas. The tests themselves are a significant expense. Yet if caught early, before the onset of its worst symptoms, SOS can be treated much more cost-effectively. This could potentially save your firm a large amount of money down the road. You are charged with making a policy recommendation to senior management.

Part A We know that SOS afflicts roughly 1 Texan out of every 1000, and let's assume that your 10 million clients are a representative sample of all Texans. No medical test is perfect, but this one is reasonably accurate: it gives a positive result for 95% of people who have SOS, and a negative result for 99% of people who do not have SOS. What is the probability that a patient has SOS, given that he or she tests positive for the disease?

Part B Is the proposed policy of free, universal SOS testing a smart financial move for your company? Use the following facts about the test/treatment costs for SOS to help you decide.

- Each test for SOS costs your company $10.
- Each false positive—that is, where the test gives a positive result for someone who does not actually have the disease—costs your company $50 in follow-up costs associated with discovering that the initial test was wrong.
- On average, each correctly identified case of SOS saves your company $10,000 in net future medical expenses.[1]

Hint: you may find it useful to recapitulate the steps in Part A for the people who test negative.

Part C In Part B, how much larger or smaller than $10,000 would the savings have to be from a correct identification of SOS in order for the proposed policy to be a "break-even" proposition?

[1] For you accounting majors, assume this is adjusted to net present value.

Problem 2: correlated random variables, linear combinations

(A) Warm-up. Suppose that X and Y are two random variables, and that their joint distribution $P(X, Y)$ has these moments:

$$\begin{aligned} E(X) &= 10 \\ E(Y) &= 20 \\ \mathrm{var}(X) &= 36 \\ \mathrm{var}(Y) &= 81 \\ \mathrm{cov}(X, Y) &= 25 \end{aligned}$$

Compute the quantities $\mathrm{E}(0.5X + 0.5Y)$ and $\mathrm{var}(0.5X + 0.5Y)$.

(B) Stocks. Suppose that you are trying to decide how much money from your investment portfolio to allocate to two stocks: US Steel, a steel manufacturer; and Alleghany Corporation, a medium-sized insurance company. Both of these companies are traded on the New York Stock Exchange. Fortuitously, they have ticker symbols X (for US Steel) and Y (for Alleghany), which exactly matches our notation for random variables. In honor of this coincidence, we will let X denote the value of $1 of US Steel stock, bought today and sold 12 months from now. Similarly, we will let Y denote the value of $1 of Alleghany stock, bought today and sold 12 months from now. For example, if $X = 0.91$ and $Y = 1.12$, then $1 of US Steel stock bought today would have lost 9% after a year, and $1 of Alleghany stock would have gained 12%.

In the file "stocksjointdist.csv", you are given a table describing a joint probability distribution for X and Y. This joint distribution is **not** just a summary of past results; rather, it represents your uncertainty about what will happen to the two stocks over the next 12 months. To simplify things, we have assumed that X and Y can only take on values between 0.76 and 1.30, in multiples of 0.03. (This is called a "discrete approximation" to a continuous variable.) The full table has 361 rows, not counting the header row, and is thus too big to fit on this sheet. But below is an excerpt:

x	y	$P(X = x, Y = y)$
0.76	0.76	3e-08
0.76	0.79	8e-08
0.76	0.82	2.1e-07
⋮	⋮	⋮
0.91	0.94	0.00171264
0.91	0.97	0.00247870
0.91	1.00	0.00333577
⋮	⋮	⋮
1.30	1.24	3.7e-07
1.30	1.27	1.5e-07
1.30	1.30	6e-08

The first column lists the possible values of X; the second column lists the possible values for Y; the third column gives the joint probability that $X = x$ and $Y = y$ for the particular values of x and y on that row. (Recall that the notation "3e-08" is shorthand for 3×10^{-8}, or 0.00000003.)

Suppose that you have w dollars to invest in these two stocks. You decide to form a portfolio by placing pw dollars in US Steel, and $(1 - p)w$ dollars in Alleghany, where $0 \leq p \leq 1$. Let W denote the total value of your portfolio after 12 months.

(i.) In terms of p and w, what linear combination of X and Y corresponds to the random variable W?

(ii.) Write down expressions for $E(W)$ and $\text{var}(W)$ in terms of p, w, and quantities you compute from the joint distribution of X and Y.

(iii.) What choice of p will maximize your portfolio's expected return? What are the expected return and standard deviation of this portfolio?

(iv.) What choice of p will minimize the variance of your portfolio's future performance? What are the expected return and standard deviation (square root of the variance) of this portfolio, expressed in terms of w?

Note: you will probably find it easiest to use a spreadsheet for this problem, although it can certainly be accomplished in R.

Problem 3: decision trees

You are the owner of a small electronics company. In 6 months, a proposal is due for an electronic timing system for downhill skiing events in the 2018 Winter Olympic Games. For several years, your company has been developing a new form of "finish beam" that records the time at which a ski racer crosses the finish line. A finish beam is a critical component in any ski timing system. They must be accurate to the millisecond and immune to poor weather. They must also be able to distinguish between a skiier and a puff of snow crossing the beam, and they must maintain a wireless data link with the starting gate in conditions which make radio transmission difficult (e.g. high mountains). When completed, yours will be superior to anything currently on the market. But your progress in research and development has been slower than expected, and you are unsure about whether your staff can perfect the technology and produce the components in time to submit a proposal. You must make a decision about whether to press ahead with an accelerated schedule of research and development, taking into account your uncertainty about future events. Consider these facts:

- The Olympic contract pays $1 million to the winner.
- If your R&D effort succeeds in developing the new beam (which will happen with probability p_1), there is a good chance (probability p_2) that your company will win the contract.
- If you decide to accelerate the R&D schedule for the purposes of developing the new beam, you will need to invest $200,000 in extra funds. With probability $1 - p_1$, this effort will fail to develop the new beam in time.
- If you fail to develop the new beam, there is a small chance (probability p_3) that you will still be able to win the same contract with your older model of timing system that has already been developed. This involves no extra cost for research and development.
- Making a proposal requires developing a prototype timing system at an additional production cost, above and beyond R&D costs. This additional cost would be $50,000 for the new model of timing system, and $40,000 for the old model.
- Finally, if you win the contract, the finished product—whether the new or the old model—will cost an additional $250,000 to produce, above and beyond the cost of building a prototype.

The following page shows a decision tree for this problem. Each oval represents the outcome of a random event; this is called a *stochastic node*. Each rectangle represents a decision that you must make under a particular set of circumstances; this is called a *decision node*. These nodes are labeled with the letters A through I. Each terminal arrow of the tree corresponds to a particular set of outcomes and decisions. These are labelled 1 through 11. This problem, along with the rules at right, will walk you through the steps of solving a decision-tree problem.

There are four basic rules for solving a decision problem using a tree:

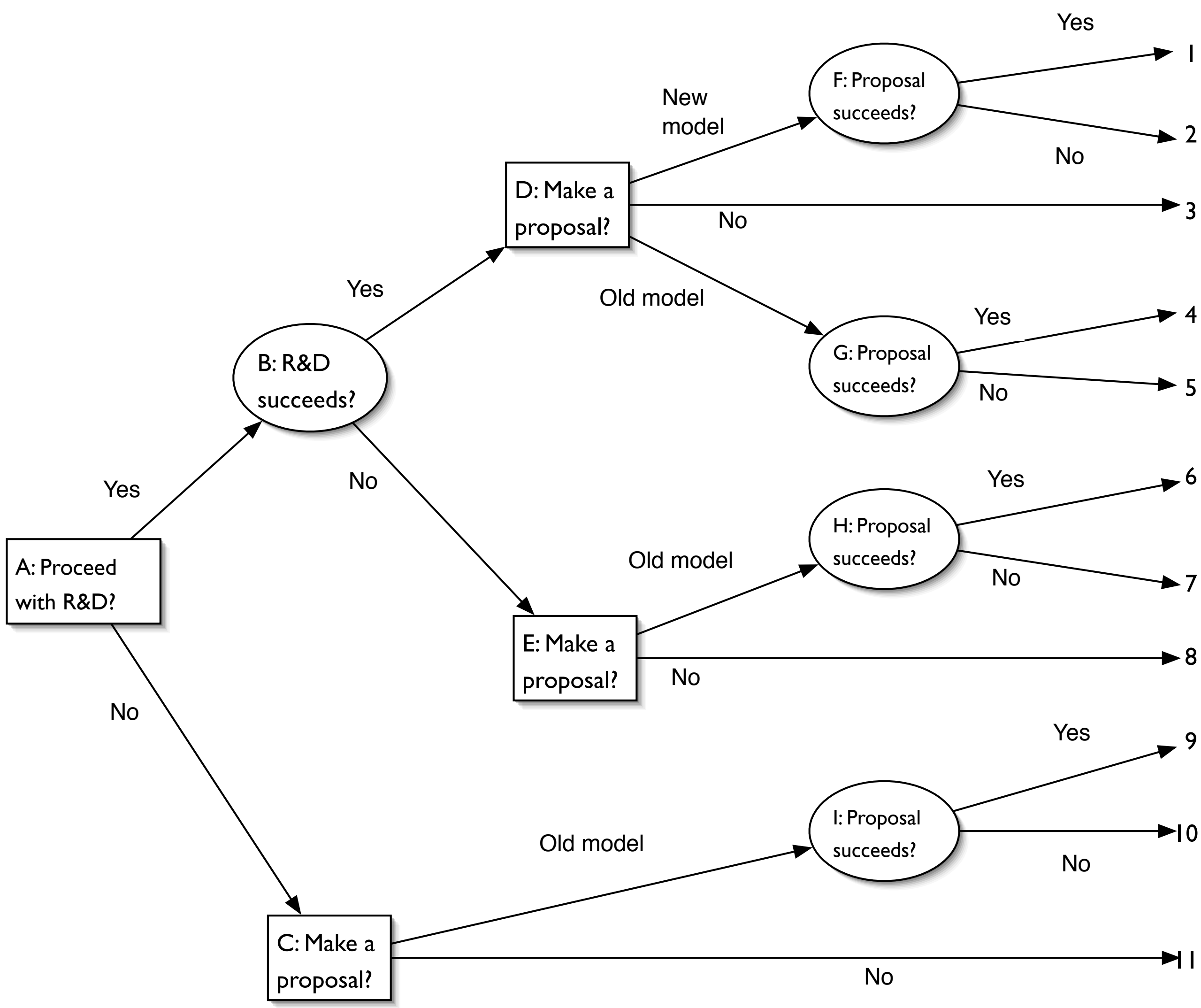
1. Ensure that the tree includes all relevant decision and stochastic nodes, in temporal order from left to right.
2. Work *right to left*, i.e. backwards in time.
3. For every stochastic node, compute the expected value of the random outcome. This quantity becomes the value of that node.
4. For every decision node, choose the course of action with the highest expected value. This quantity becomes the value of that node.

Part A: Each of the 11 terminal arrows of the tree, corresponds to a particular pattern of events that may or may not happen. For each node, fill in the amount of net profit you would make if those events actually happen. Make sure to account for all costs along each path of the tree.

Part B: For the right-most four stochastic nodes of the tree (F, G, H, and I), fill in the probabilities of the possible outcomes (two for each node), and then compute the expected value of the node in terms of p_1, p_2, and p_3. Remember, the outcome at that node is a random variable, and we know how to calculate expected values of random variables.

Part C: For decision nodes D, E, and C, write down a *decision rule*—that is, a criterion expressed in terms of the relevant probabilities (p_1, p_2, and p_3) that tells you which decision you should make to maximize utility in that set of circumstances. Derive these rules by hand, and show your work. You may use software to help with routine calculations, but not to derive the rules.

Part D: Suppose $p_2 = 0.8$ and $p_3 = 0.1$. Write down an expression for the expected value of Node B in terms of p_1. What value of p_1 makes you indifferent between proceeding with the accelerated R&D effort, or not?


A: Proceed with R&D?
Yes
No
B: R&D succeeds?
Yes
No
C: Make a proposal?
D: Make a proposal?
E: Make a proposal?
New model
No
Old model
F: Proposal succeeds?
G: Proposal succeeds?
H: Proposal succeeds?
I: Proposal succeeds?
Yes
No
1
2
3
4
5
6
7
8
9
10
11