Exercises 8 · Bayes's rule; probability models

Due Monday, April 24, 2017

Problem 1: updating conditional probabilities

Suppose that you work in the analytics office for one of the state's largest health-insurance companies. Your first assignment is to study the cost-effectiveness of instituting a universal test for a disease called SOS. Your firm is thinking about making the test free and universal 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 understanding the properties of the test.

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. In light of these numbers, what is the probability that a patient has SOS, given that he or she tests positive for the disease?

Problem 2: sequential betting

Suppose you have w_0 dollars to your name, and that you are offered the following bet.

- With probability $p = 0.52$, you will win the bet.
- With probability $1 - p = 0.48$, you will lose the bet.

It sounds like a good bet: you are more likely to win than to lose. Moreover, you get to choose what fraction c of your total wealth w_0 to wager on the outcome of the bet.

(A) Suppose that your current wealth is $w_0 = \$1000$, and you decide to risk $c = 0.10$ (i.e. 10%) of your wealth on this bet. What is your expected wealth, w_1, after the outcome of the bet becomes known?

(B) Imagine now that you get to repeat the bet as many times as you want. After every single round of betting, you decide to risk the same fraction $c = 0.1$ of your current wealth on the next bet. In other words, if you have w_t dollars after round t of betting, you place $0.1w_t$ dollars on the next round's wager. If you win, you will wealth will be $w_{t+1} = (1+c) \cdot w_t$. If you lose, it will be $w_{t+1} = (1-c) \cdot w_t$.

Suppose you start with \$1000. Simulate 10,000 rounds of this bet.[1] (I recommend that you build on our R scripts from the walk-through on Monte Carlo for sequential events. But you can also make a spreadsheet with 10,000 rows.) Make a plots of your simulated trajectory of wealth w_t over every round from $t = 1$ to $t = 10000$ (the betting round t should be on the x-axis). What happens after 10,000 rounds of betting? Are you rich or broke? (Repeat the simulation several times if you need to, in order to convince yourself of what happens here.) In light of the answer from Part A, do you find this surprising?

[1] Why 10,000? Because that's roughly the number of trading days over a 40-year period of investment.

(C) Now repeat the simulation—except this time, only risk 0.5% of your current wealth (that is, $c = 0.005$, or 1 part in 200) at every round of betting. As before, plot your simulated value of current wealth w_t at every step from 1 to 10,000. This time, what happens after 10,000 rounds? Are you rich or broke?

(D) Experiment with your Monte Carlo simulation to find a value of c that you like best in order to maximize the long-term growth of your portfolio. Comment on the wisdom of the following: "If your goal is to ensure the long-term growth of your capital, you should make bets that carry the highest possible expected return."

Problem 3: a better model for soccer

Go read the article "One match to go!" by Spiegelhalter and Ng, available here: `http://faculty.chicagobooth.edu/nicholas.polson/teaching/41000/speigelhalter-epl.pdf`. They describe how they formulated an approach for predicting the probability of different outcomes for soccer matches. It is better than the simple approach we took in the course packet (though probably not as good as what actual bookies use).

Now go get data from this year's English Premiere League soccer season. For example, you can certainly find it here: `http://www.soccerstats.com/latest.asp?league=england`. You can get home/away splits by clicking on "Home/Away" under the "Statistics" button. Replicate Spiegelhalter and Ng's approach using this year's data. (This is probably easiest to do in Excel, although you can certainly use R if you want.) What is your estimated probability distribution of likely results for the match on April 25th between Chelsea and Southampton? How about the match on April 30th between Tottenham and Arsenal?