Recitation 9

In Recitation 7, you computed the optimal betting strategy for a completed NCAA tournament. In this week's recitation, you will apply (*stochastic*) dynamic programming techniques to select a bet in the same style pool for a tournament that has not yet taken place. To start, you will use data that was prepared just prior to the 2006 NCAA men's basketball tournament.

Name and NetID:
Section:

1 Introduction

Once again, 64 teams compete in asingle elimination bracket consisting of 6 rounds. In each game, the winner advance to round 2, 16 to round 3, 8 to round 4, etc., until the finals, in which only two teams compete for the championship.) So, there are 63 games in total, and a team must win 6 games in a row to become the champion. The bracket is split into four regions of 16 teams apiece, and each team is given a seed from 1 to 16 (where this "seed"ing is a hypothesized ranking of these 16 teams from best to worst). The teams that have played the best during the season are rewarded with the lower numbered seeds, so the four # 1 seeds are the favorites to win, while the four # 16 seeds are perceived to have little chance of winning.

The pool was invented by Robin Lock and popularized by former Cornell ORIE faculty member Rick Cleary. The teams are each given a price based on their seed, according to the table below. The costs are given in cents.

seed	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
cost	25	21	18	15	12	10	8	6	5	4	3	2	1	1	1	1

Before the start of the tournament, each player who enters the pool is allowed to buy up to one dollar's worth of teams. For instance, one could buy all four # 1 seeds for 25 cents apiece. Or one could buy three # 1 seeds, one # 3 seed, one # 11 seed, and all four # 16 seeds. At the end of the tournament, each player adds up the total number of games won by teams she purchased, and the player with the highest total wins the pool.

In this probabilistic setting, the aim will be to place a bet to maximize the ttoal expected number of games that teams selected win. Recall that expectation has an absolutely critical "linearity property". For any given bet, we have selected a subset S of the teams that we are hoping will do well. IF W_k is a random variable that specifies the number of games that team k wins, then we wish to choose a feasible subset S such that the expected value of $\sum_{k \in S} W_k$ is maximized. The linearity of expectation tell us that we need only compute $E_x[W_k]$ and then sum these values in order to evaluate the objective function value for the selection S.

The first step is to evaluate the expectations $E_x[W_k]$. First of all, what is the input that are given? We are given, for each pair of teams j and k, the probability p_{jk} that, if there were a game between these two teams, team j would beat team k. (So, $p_{kj} = 1 - o_{jk}$.) Furthermore, we will assume if these two teams meet in the tournament, then p_{jk} specifies the probability of team j winning, independent of the possible outcomes in the earlier rounds that led to this pairing.

2 Dynamic programming formulation

The following steps will guide your dynamic programming solution for this problem.

1. Of course, how do we even know what values to use of p_{jk} for the upcoming tournament? One way is as follows: suppose we know the odds, for each team k, that is currently given by Las Vegas bookmakers, that it wins the tournament. That is, we have an estimate q_k of the probability that team k will win the entire tournament. We will impute the value p_{jk} to be $\frac{q_j}{q_j+q_k}$. Why does this makes sense?

2. In the data file posted on blackboard, there is an advance prediction for the probabilities q_k for last year's tournament (prior to its start). Take a look at the data file.

   Let P_{kl} be the probability that team k reaches round $l+1$ of the tournament (or in other words, the probability that team k wins at least l games). How doe $E_x[W_k]$ and $\sum_{l=1}^{6} P_{kl}$ relate?

3. What do you know about P_{k0} for each team k?

   For each team K, and each round l, there is a set of teams that compete to play against team k in round l, when team k has successfully won its first $l-1$ games; let S_{kl} denote that set of teams. (SO, for $L=1$, this set has exactly one member, the team that k plays in round 1.)

   Explain why
   $$P_{k,l+1} = P_{k,l} \cdot \sum_{j \in S_{kl}} p_{kj} \cdot P_{jt}.$$

4. You have just created the basics of a dynamic program to compute the value P_{kl}, for each team K, and each $L = 0, \ldots, 6$. You know the values for $l = 0$. Given these and the matrix $(p_{jk}$, you can compute the values P_{k1}. Next, compute P_{k2}, and so forth, until you compute P_{k6} for each team k. And now you can compute $E_x[W_k]$. We shall let this expectation be denoted E_k.

   Next, you will construct an AMPL file that takes (p_{jk}) values, and the sets S_{kl}, and computes the values E_k. You will see on blackboard, a shell of a mod file, and a sample data file.

   Note that the sets S_{kl} are not given explicitly, but instead computed by their own recursion. Spend a few minutes to try to figure out what this computation is doing.

   Complete the mod file and use it to compute the values E_k. In addition to displaying these values, you should also display `sum k in teams E[k]`. This sum has a value that should be easy to predict. Even if you cannot figure this out in advance, can you explain, after you see it, why this value is correct?

5. Now consider the optimization problem of choosing the right bet. You want to maximize over all feasible choices, the maximum expected total number of games won. It is easy to see that this is now exactly as what you had to do last time, except that instead of being given the number of wins that each team k did in fact win, you are given the expected number of games won by team k (relying again on the linearity of expectation). Incorporate last week's model into the AMPL file that you have constructed to compute the values E_k, to compute the optimal bet for the sample data file. THere is a sample mod file posted for the knapsack calculation that you have to do in this part of the exercise that can be added to your computed mod file to compute the optimal bet.

   What is the optimal number of expected games that you can attain by a bet of at most 100 cents for the given data for last yeaer's tournament?

6. FUrthermore, there is an AMPL script on blackboard that iterates through the 64 teams to compute the optimal solution. Use this script to obtain your optimal bet. On which teams does this model say that you should bet?