MS&E 235: Lecture 6
Prof. Ashish Goel
Apr 16, 2009
Notes by Rio Goodman

Lecture 6: More Uncertainty

In this lecture we will see more on analyzing a decision under uncertainty. We will use the multi-armed bandit problem and the Gittins index from last lecture.

Uncertainty and Markets

Uncertainty and markets are two sides of the same coin. It is important to understand both the uncertainty and the market. Consider two types of treasury bills: Type 1 has a reward of \$2 in the second year, Type 2 has a reward of \$1.5 in the second year. Clearly, you would choose Type 1. But what if there is some uncertainty in the market? That is when the problem becomes non-trivial.

Uncertainty is especially important in online markets because there is large scale data available to make individual decisions and often there is little statistical information available due to the long tail. We will learn more about long tailed distributions later.

Gittins Index

Imagine that we are picking a basketball team from the students in our class. We would like to choose the tallest students. In such a situation it is obvious what to do. We simply measure everyone and compare the measured heights. But in a mutli-armed bandit situation, it is not as obvious. How do we compare different arms with different histories? The problem is that we have no “height” as we did for measuring students.

Example

We have two possible arms, A and B, with current success probabilities of 0.4 and 0.6 respectively. We want to compare the arms as we compared the heights of the students, so that we can determine the optimal arm to pull. The arms have success trees:

A: 0.4 (children: 0.15, 0.65)

B: 0.6 (children: 0.5, 0.7)

The first tree tells us that the current success probability associated with A is 0.4. If A is pulled and experiences a success, then its probability increases to 0.65, but if it experiences a failure, it drops to 0.15. Let us assume that for the first pull, the success probability for each arm is 50%. If we are only given one opportunity to pull an arm, and then, given that experiment, must decide which arm to pull thereafter, our decision will be arrived at like this:

Case 1: If we choose A, and it is a failure, then $0.15 < 0.6$ so choose B.
Case 2: If we choose A and it is a success, then $0.65 > 0.6$ so choose A.
Case 3: If we choose B, and it is a failure, then $0.5 < 0.4$ so choose B.
Case 4: If we choose B and it is a success, then $0.7 > 0.4$ so choose B.

If we choose A, then with probability 0.5, we will be in the Case 1, and with the same probability we will be in Case 2. So the expected profit will be: $E[\text{Profit}|A] = 0.5 \times 0.6 + 0.5 \times 0.65 = 0.625$. In other words, if

we choose to experiment with A, our expected reward is 0.625.
Similarly, $E[\text{Profit}|B] = 0.5 \times 0.5 + 0.5 \times 0.7 = 0.6$

$\Rightarrow E[\text{Profit}|A] > E[\text{Profit}|B] \Rightarrow$ Play arm A.

In our basketball game, this might correspond to two students, A and B, where student A is taller than student B. In this example, we were able to get around the fact that we have no notion of "height" in the two-armed bandit problem by computing the expected values of the probabilities. But now suppose we have another student, C, who is the same height as B. We would still choose student A because clearly, if A is taller than B, and C is the same height as B, then A is also taller than C. But we will see that this is not the case with the multi-armed bandit problem.

With C added to our list of options, playing A will result in the same expected profit. But playing B will result in a different expected profit, because if B turns out to be 0.5, then we can choose C. $E[\text{Profit}|B] = 0.5 \times 0.6 + 0.5 \times 0.7 = 0.65 \Rightarrow E[\text{Profit}|B] > E[\text{Profit}|A] \Rightarrow$ Play arm B (or C). Our answer changed by introducing a new arm, C.

This example illustrates that with uncertainty, unlike height, the optimal choice depends on what other options are available. In other words, an individual attribute, such as height, cannot necessarily be used to compare different options. Just because one arm may have a higher expected profit, does not mean it is the optimal choice. This is where the Gittins index comes in.

The Gittins Index Theorem tells us that, if our goal is to maximize total discounted reward, then there exists a function, $g(\alpha, \beta, \theta)$, such that it is always an optimal strategy to play the machine i with the highest $g(\alpha_i, \beta_i, \theta)$. That is why the Gittins Index Theorem is so powerful in this setting.

So the Gittins Index is a notion similar to "height." R_p gives us our scale. And finally, V gives us a method for testing whether a point is higher or lower than our height, thus allowing us to hone in on a value.

As we saw in the last lecture, we will have a value function, V, given by:

$$V(p; \alpha, \beta, \theta) = max\left\{\frac{p}{1-\theta}, \frac{\alpha}{\alpha+\beta} + \theta\left(\frac{\alpha}{\alpha+\beta}V(p; \alpha+1, \beta, \theta) + \frac{\beta}{\alpha+\beta}V(p; \alpha, \beta+1, \theta)\right)\right\}$$

Where p is the success probability of a reference machine, α and β are the historical number of successes and failures of the machine being considered, and θ is the discount factor. We change our reference machine to find the point where we are indifferent between playing the reference machine and playing the machine we are considering. We do this by finding the lowest p such that $V(p; \alpha, \beta, \theta) = \frac{p}{1-\theta}$. This gives us our Gittins index.

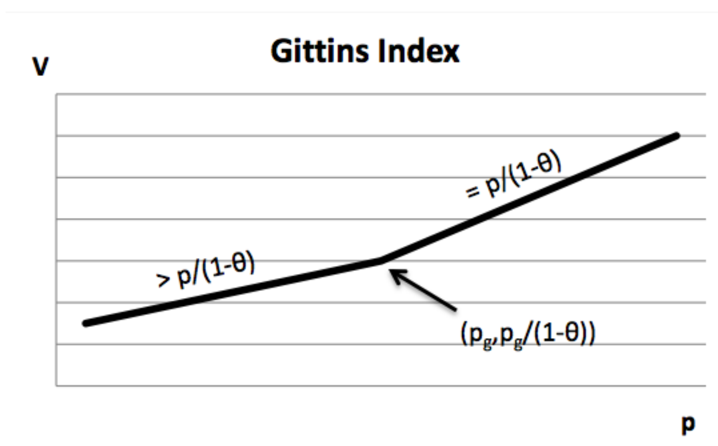

Figure 1: Gittins Index

If we have a machine, (α, β), we try to find a Gittins index p_g. In the graph above, V is on the y-axis, and p is on the x-axis. At a certain point, p_g, we will have the R_p machine and the (α, β) machine equally as desirable, with the value, $V = \frac{p_g}{1-\theta}$. Anywhere to the right of p_g, playing the R_p machine will yield a higher expected reward ($V = \frac{p}{1-\theta}$), and anywhere to the left of p_g, playing the (α, β) machine will yield a higher expected reward ($V > \frac{p}{1-\theta}$).

In order to actually calculate the Gittins index, we can use a spreadsheet with formulas. The cells will contain the values of V, the rows will be β values, and the columns α values. This is provided in an excel file.