# MS&E 235, Internet Commerce

## Stanford University, Winter 2007-08

**Instructor: Prof. Ashish Goel, Notes Scribed by Shrikrishna Shrin**

# Lecture 4: Decision Making Under Uncertainty

### A sample experiment

Consider the experiment of displaying an advertisement. The actions possible are either a click or no click. Consider two variables $\alpha$ and $\beta$. If the ad is clicked, let us increment $\alpha$ and if it is not, $\beta$. Let us also assume that the advertiser is risk neutral.

Given the above information, we can estimate the probability of the ad being clicked when it is displayed next to be $\alpha/(\alpha+\beta)$ and that of the ad not being clicked to be $\beta/(\alpha+\beta)$. Furthermore, $(\alpha+\beta)$ gives us an idea of the confidence in the estimate.

This model captures the probability as well as the confidence at the same time. Also, the model evolves over time. $\alpha$ and $\beta$ not only dictate the probability of success but also how the model is going to evolve. $(\alpha, \beta)$ is known as a prior or a belief.

### Example

Let us consider two slot machines. One slot machine is guaranteed to result in a win or loss with probability 0.5. Let us call this machine a fair machine. You have another slot machine in which three attempts have been made out of which two resulted in a loss and one resulted in a win. If you had 1 dollar where would you invest it?

Considering probabilities of a win, the fair machine has a .5 probability of a win whereas the other machine only has a .33 probability of a win. Also, the fair machine has a greater confidence as it is known with certainty that the probability of a win is 0.5. On the other machine, however, only three trials have been made and therefore has a much lower confidence associated with the chance of a win on it.

The question of which machine to play can be explained via an explore/exploit trade off. Assume that you have infinite cash and infinite time except you can play only once everyday. With the given information, if it is optimal to play the fair machine today, it will be optimal to play the fair machine every other day as you are presented with the same problem once again. However, assuming that there is no discount factor, that is money today is worth as much as money tomorrow, it will be a better strategy to play the $(1, 2)$ machine to get more information and a higher confidence on the probability of winning. This is because there is a possibility that the probability of a win on the machine is greater than 0.5. If it becomes clear that this is not so, say there are 11

consecutive losses, we can shift to the fair machine and continue to play as you now know that the $(1, 2)$ machine is unfavorable with greater confidence.

On the contrary, if the discount factor was zero, that is, money tomorrow is worth nothing today, it is a better strategy to play the fair machine as it has a higher current probability of a win. If money tomorrow is the same as money today, it will be a better strategy to explore. If the money tomorrow is worth nothing today, it will be better to exploit. In general, there is a trade off between exploit and explore (when the discount factor is between zero and one).

Assume that there are $N$ slot machine arms (or $N$ advertisements).

Every arm $i$ has ($\alpha_i$,$\beta_i$) associated with it. There is a discount factor $\theta$ between $(0, 1)$ ($\theta$ is how much a dollar tomorrow is worth today). If $\theta = 1$, play the machine for which less information is available. If $\theta = 0$, play the machine that has a higher current probability of a win. An interesting case is when $\theta$ is between 0 and 1. In such a case, the strategy is to maximize the total expected discounted reward. Such a problem is also called the multi-arm bandit problem.

**Gittin's Theorem**

There exists a function $g(\alpha_i, \beta_i, \theta_i)$ such that it is an optimum strategy to play the arm with the highest value of $g(\alpha_i, \beta_i, \theta_i)$. The beauty of this theorem is that the value of the gittin's index for each machine can be calculated independent of other machines.