**OPNS530 Sequential Learning** Lecture 1 - March 28th 2017

# Introduction

*Lecturer: Daniel Russo* *Scribe: Alexej Proskynitopoulos, Yanir Cohen*

## Overview

Assignments:

- Lecture notes scribing
- Occasional HW
- Course Project

The course project could be a literature review, an implementation project or on an application of the methods developed and discussed in class. The course will contain several modules:

### Module 1: Sequential/Active Hypothesis Testing

Suppose you want to hire a freelance programmer. You design a questionnaire to assess candidate capabilities to decide if one is a good fit. Assume your test consists of $n$ attributes. Consider a vector $\theta = (\theta_1, \ldots, \theta_n)$ of 'skills' of a programmer (e.g. "Java", or "Machine Learning", etc.). The programmer is then asked a question, $q$. We observe its answer. Denote his answer with $q = 1$ if it is correct and 0 otherwise. The probability of him answering correct is

$$prob = logistic(\theta^\top q) = \frac{e^{\theta^\top q}}{1 + e^{\theta^\top q}}.$$

Consider now the following hypothesis:

$$H_0 : \theta \in \Theta_0, \qquad H_1 : \theta \in \Theta_1,$$

where $\Theta_0$ (the set of skill sets for which we would hire the programmer) and $H_1$ (the set of skills for which we would not) are disjoint. We are interested in defining an effective test and would like to know (1) if a candidate is a good programmer ? how much time it would take to define he/she is good? and (2) how many questions I should ask to decide that someone is not a good fit.

### Module 2: Bandit Learning

Consider the following example of an online shortest path problem. Everyday we send a truck from a source $S$ to a destination $D$ to deliver goods. As we send it, we observe some feedback on how long it took the driver to deliver the items. The driver needs to visit the destination via a path of a weighted graph.
Our goal is to minimize

$$\mathbb{E}[\text{total travel time over many trips}].$$

We cannot try every path as typically the number of those is very(!) large, possibly exponentially. Instead, we would like to work in a simple algorithmic framework to obtain theoretical guarantees of our algorithms. We compare our performance against the performance we would achieve by knowing the best path (this is called the regret). Typically, our algorithms address the trade off between exploration and exploitation.

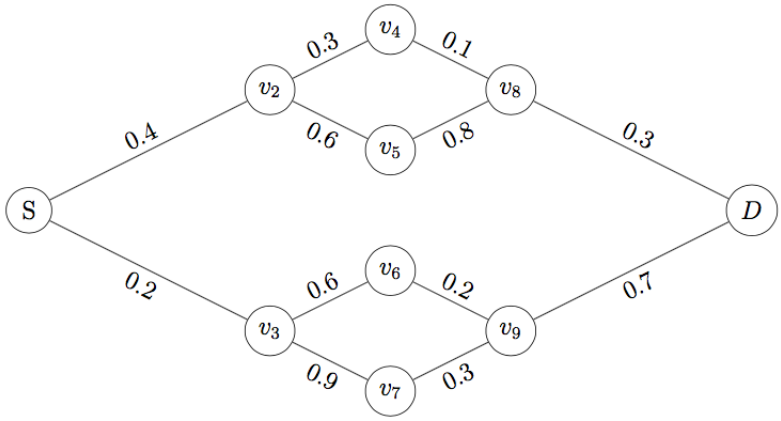


**Figure 1**: A shortest path problem with ten vertices.

### Module 3: Reinforcement Learning

We learn to optimize an Markov decision process (MDP) from observed actions and state-transitions and rewards. Consider the following example: patients arrive at a hospital. Each patient comes in a state, captured, for example, by his weight, age and several other characteristics. He is then exposed to a treatment that changes his state and gives a reward, based on his health improvement (or deterioration). We then expose him to another treatment and observe his new state and so on for a couple of turns, before we move on to the next patient. Note that this is a case of delayed consequence. It is not the traditional bandit problem; it would take time to fully observe the consequences of a past decision.

## 1 Sequential Hypothesis Testing

The topic of sequential hypothesis testing dates back to Wald (1945). Say we observe a random variable $X_i$ at time $i$. We are interested in a binary hypothesis test of:

$$H_0 : X_1, X_2, \dots \overset{i.i.d.}{\sim} f_0, \quad vs \quad H_1 : X_1, X_2, \dots \overset{i.i.d.}{\sim} f_1,$$

where $f_0, f_1$ are distribution functions. Assume that we have no structural constraints so that we can continue to observe data indefinitely. Let $\theta \in \{0, 1\}$ denote the true unknown true parameter (so $X_i \sim f_\theta$). Initially, we observe $X_1$ at which point we either stop and declare $\theta = 0$ or $\theta = 1$, or we continue and observe $X_2$. After observing $X_2$ we can again either declare $\theta$ or continue to observe data and so on. This gives rise to the decision tree in figure 1.
Rigorously, a sequential test[1] is a sequence of functions $\Psi = (\psi_1, \psi_2, \dots)$ where

$$\psi_n : (X_1, \dots, X_n) \to \begin{cases} \text{stop and return } H_0 \\ \text{stop and return } H_1 \\ \text{continue.} \end{cases}$$

Let $(T, D)$ be a pair of a (random) stopping time and testing decision. We are typically interested in quantities such as $\mathbb{E}_0[T] = \mathbb{E}[T|H_0]$, $\mathbb{E}_1[T] = \mathbb{E}[T|H_1]$ as well as $\mathbb{P}_0(D \neq 0) = \mathbb{P}(D \neq 0|H_0)$ and $\mathbb{P}_1(D \neq 0) = \mathbb{P}(D \neq 0|H_1)$ . Our goal will be to minimize these quantities. Clearly there is (in general) no way to minimize all four of these simultaneously.

[1] In this class we will occasionally refer to a sequential test as a rule, procedure or algorithm, and we will use these term somewhat interchangeably.

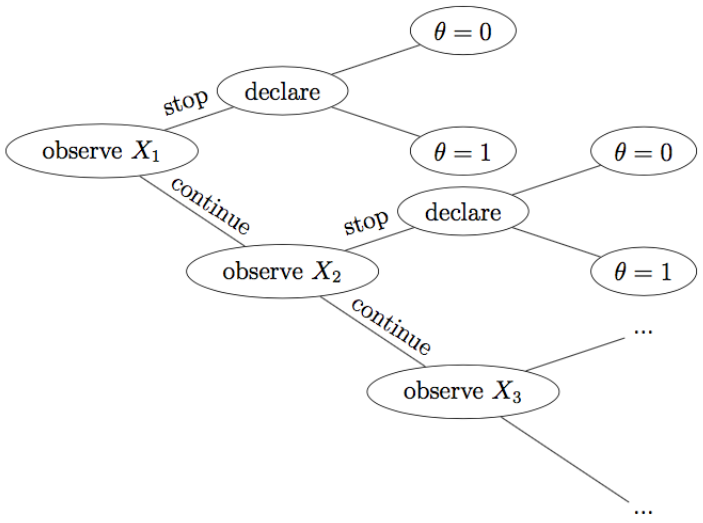


**Figure 2**: Decision three corresponding to the sequential test.

**Example 1** (Benefits of sequential test, a contrived example)**.** *Say*

$$H_0 : X_1, X_2, \cdots \sim Ber(1), \qquad H_1 : X_1, X_2, \cdots \sim Ber(1/2).$$

*We wish to guarantee* $\mathbb{P}_0(D \neq 0) = 0$ *as well as* $\mathbb{P}_1(D \neq 1) \leq \delta$ *for some small* $\delta \geq 0$*. How many samples are required? Say* $T = n$*. Then*

$$D = \begin{cases} 0 & \text{if } X_1 = X_2 = \cdots = X_n = 1 \\ 1 & \text{otherwise.} \end{cases}$$

*and* $\mathbb{P}_1(D \neq 1) = (1/2)^n \leq \delta$ *provided that* $n \geq \log_2(1/\delta)$*. The sequential test will stop if you see a 0 or after* $\log_2(1/\delta)$ *turns. It is easy to see that* $\mathbb{E}_0[T] = \log_2(1/\delta)$*. Under* $H_1$*,* $T \sim Geo(1/2)$ *and therefore* $\mathbb{E}_1[T] = 2$*.*

The (Bayesian) motivation for the sequential test is as follows: say

$$\theta = \begin{cases} 1 & \text{prob} = \pi_0 \\ 0 & \text{prob} = 1 - \pi_0, \end{cases}$$

and that $X_1, X_2, \ldots | \theta \sim f_0$. The posterior distribution is

$$\pi_n := \mathbb{P}(\theta = 1 | X_1, \ldots, X_n) = \frac{\pi_0 \prod_{k=1}^n f_1(x_k)}{\pi_0 \prod_{k=1}^n f_1(x_k) + (1 - \pi_0) \prod_{k=1}^n f_0(x_k)},$$

and

$$1 - \pi_n = \mathbb{P}_0(\theta = 0 | X_1, \ldots, X_n).$$

Say we stop if

$$\pi_n \geq A \iff \frac{\pi_n}{1 - \pi_n} > \frac{A}{1 - A} \iff \frac{\pi_0 \prod_{k=1}^n f_1(x_k)}{(1 - \pi_0) \prod_{k=1}^n f_0(x_k)} \geq \frac{A}{1 - A}.$$

Let $\Lambda_n := \prod_{k=1}^n f_1(x_k)/f_0(x_k)$ be the likelihood ratio. If $\pi_n \geq A$, then $\Lambda_n \geq A/(1 - A)$. Our sequential probability ratio test (SPRT) is then:

$$\begin{cases} \text{stop if } \Lambda_n \notin [A, B] \\ \text{continue otherwise.} \end{cases}$$

For some thresholds $A$ and $B$. Questions:

1. Is this a valid test?
2. How many samples are required?
3. Is this optimal?

Optimality results: "Ignoring" overshoots, the SPRT with $(A, B) = function(\alpha, \beta)$ minimizes both $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ among all rules under which

$$\mathbb{P}_0(D \neq 0) < \alpha, \qquad \mathbb{P}_1(D \neq 1) < \beta.$$

Notice that

$$\log \Lambda_n = \sum_{k=1}^{n} \log \frac{f_1(x_k)}{f_0(x_k)} = \sum_{k=1}^{n} Z_k,$$

for $Z_k = f_1(x_k)/f_0(x_k)$. Consider the test $\log \Lambda_n \notin [\log A, \log B]$. It is easy to see that $\log \Lambda = \{(\log \Lambda_n)\}_{n \in \mathbb{N}}$ is a random walk with drift. For $n$ large we expect from the CLT that $\log \Lambda_n \approx n\mathbb{E}[Z_1] + O(\sqrt{n})$.

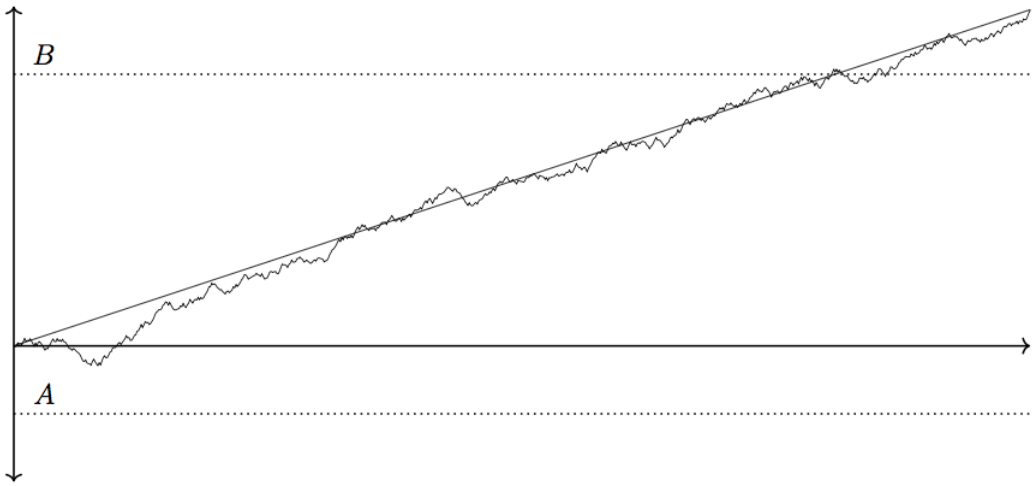


**Figure 3**: Log likelihood with positive drift and decision interval $[A, B]$.

Assuming an upward drift, we cross the boundary at $B$ when

$$n\mathbb{E}_1[Z_1] \geq \log B \Rightarrow n \approx \frac{\log B}{\mathbb{E}[Z_1]}.$$

Similarly

$$\mathbb{E}_0[T] \approx \frac{\log A}{\mathbb{E}_0[Z_1]}.$$

**Remark** $\mathbb{E}_1[Z]$ is the Kullback-Leibler divergence between $f_1$ and $f_0$, also denoted as $KL(f_1||f_0)$.