Lab #10: Bayesian Search Theory

Due: Tuesday, December 1, 2015, in class

Finding the USS Scorpion

In late May, 1968, the U.S. Navy submarine USS Scorpion failed to arrive at its naval base in Norfolk, Virginia. In order to locate the vessel, the U.S. Navy employed "Bayesian search theory", which offers a more efficient strategy then simply combing the vast ocean depths. Details of the search for the USS Scorpion are reported in:

- H.R. Richardson and L.D. Stone (1971). Operations analysis during the underwater search for Scorpion. *Naval Research Logistics Quarterly*, vol. 18(2), pp. 141–157.

Bayesian search theory had previously been used to successfully recover a lost hydrogen bomb, and has since been used to find many lost objects, including the "black boxes" of Air France flight 447 that went down in the Atlantic in 2009.

Your assignment for this lab will be to re-enact the search for the USS Scorpion via simulation in Matlab. You will first derive the equations needed to conduct the Bayesian search, and then implement these equations in a script that will iterate through phases of the search.

Bayesian Search Theory

First, the search area can be broken up into a grid of cells:

$i=1$	2	3	4	5
6	7	$\cdots$		

Let's define some things:

$$i = \text{indices of grid cells}$$
$$Y_i = \text{sub's true existence in } i^{th} \text{ cell (1 means it is there)}$$
$$\pi_i = \text{probability of occurrence} = Pr(Y_i = 1)$$

However (and this is a critical component to Bayesian search theory), the searchers may not find the submarine even if it is in the cell being searched. Because of the limitations in technology for scanning the ocean floor, there is some "probability of detection" of the submarine. So:

$$\begin{aligned} Z_i &= \text{search results in the } i^{th} \text{ cell (1 means it is detected)} \\ p_i &= \text{probability of detection} = Pr(Z_i = 1 | Y_i = 1) \end{aligned}$$

The probabilities p_i and π_i are prior values obtained from experts before the search. The following uses Bayes' Theorem to develop equations for updating these probabilities as more evidence is obtained (i.e., cells are searched).

We can assume cells are i.i.d., and that both the detection and occurrence of the submarine follow the Bernoulli distribution (either you find it or you don't; either it's there or it isn't):

$$\begin{aligned} Z_i | Y_i &\sim Ber(Y_i p_i) = (Y_i p_i)^z (1 - Y_i p_i)^{1-z}, \quad i = 1, \ldots, n \\ Y_i &\sim Ber(\pi_i) = (\pi_i)^y (1 - \pi_i)^{1-y}, \quad i = 1, \ldots, n. \end{aligned}$$

In particular, we are interested in the case where the i^{th} cell contains the submarine ($Y_i = 1$), but we fail to detect it there ($Z_i = 0$). Use Bayes' Theorem to get the posterior probability of this case (we give you the answer here; your job is to derive it):

$$= \frac{(1 - p_i)\pi_i}{1 - p_i \pi_i}$$

We use this posterior probability to update the occurrence probability, π_i, if we search cell i and don't find the submarine:

$$\pi_{i,\text{new}} = \frac{(1 - p_i)\pi_{i,\text{old}}}{1 - p_i \pi_{i,\text{old}}}. \tag{1}$$

Here, the detection probability p_i stays constant. Also, note that $\pi_{i,\text{new}}$ is ALWAYS less than $\pi_{i,\text{old}}$ for all $0 < \pi_{i,\text{old}} < 1$.

Conversely, if the submarine is not detected in the ith cell, then our newly-acquired evidence suggests that it is in another cell. In other words, we expect that its probability of occurrence should increase in the other un-searched cells, $j \neq i$. Again, use Bayes' Theorem to arrive at an expression for updating the probability of occurrence in other cells. (Hint: you don't have the distribution for $Z_i | Y_j$, so think about what the value of this likelihood is in terms of its physical meaning).

$$= \frac{\pi_j}{1 - p_i \pi_i}$$

Thus,

$$\pi_{j,new} = \frac{\pi_{j,old}}{1 - p_i \pi_{i,old}} \tag{2}$$

where now $\pi_{j,new}$ is ALWAYS greater than $\pi_{j,old}$.

Simulating the search for the USS Scorpion

- **Goal:** Using Equations 1 and 2 that you found above, and given some initial data, develop a Matlab script that will iterate through a hypothetical search for the missing submarine.
- **Data:** In Richardson & Stone (1971), we are given that the search is conducted over a 20-cell by 20-cell grid. Even though we are searching over a 2D grid, we index the cells of the grid with only one variable (i; see figure on page 1), and therefore a vector format is appropriate for our Matlab objects.The Matlab file “Lab7.mat” contains the following four objects:
  1. Y – a vector of length 400 indicating the location of the USS Scorpion
  2. op – a vector of length 400 with the prior occurrence probabilities for each grid cell (as determined by expert opinion; see Richardson & Stone (1971) for details of how these were determined).
  3. Z – a vector of length 400, all zeros, indicating the detection of the USS Scorpion
  4. dp – a vector of length 400, all 0.46, indicating the detection probability (this is a constant value over all space and time; it depends on the technology used for sensing the submarine, and more details are provided in the article).
- **Pseudocode:** Starting with the data above, your Matlab script should:
  1. While the submarine is not found, iterate through some undefined m steps of the search process.
  2. At each step, one cell is searched. This cell is chosen as the cell with the highest occurrence probability (op).

3. Conduct the "search" via a random Bernoulli trial: binornd(1, p), where $p = dp_i \times Y_i$
4. If the submarine is not found, update the occurrence probabilities for all cells via the equations developed above.
5. If the submarine is found, stop searching.

Make sure to record how many steps, m, your search takes. Because this relies on a random search at each step, m may vary on different runs of this script, and may even be infinite (very very very unlikely, but recall that ctrl+c breaks Matlab out of infinite loops). Also make sure to record your vector of occurrence probabilities (op) at each step, as you will have to use those to answer some of the following questions.

Deliverables

Besides showing your derivation of the posterior probability equations for cells i and j above, and including your Matlab code, your lab report should answer the following questions.

1. Make maps of: a) your initial occurrence probabilities, and b) your final occurrence probabilities over the entire 20×20 grid. (Hint: use the commands vec2mat() and imagesc()). Include color bars, and include these plots in your report. In general, how have these probability values, and in particular their spatial distribution, changed over time?
2. Graph the occurrence probability in the cell containing the USS Scorpion (Y == 1) versus search steps (i.e., $1, 2, \ldots, m$). Include this graph in your report. How many searches of THIS cell did it take to find the submarine? What part(s) of this graph correspond to this cell being searched? What pattern does the history of occurrence probability have for this cell? How do the posterior probability equations relate to this pattern?
3. How is this Bayesian search method more efficient than searching the cells in order?