

Homework 2

How long until we find it?

Due: Session 6, 11:59 PM PT

*Student Name:**Instructor: John Lipor***Instructions**

Submit your work as a PDF through the course submission system. If you need an accommodation or an alternate format, contact the instructor.

Note: For all questions asked in this course, unless the question is simply asking about definitions, please justify all answers. Stating the correct answer without justification will earn no credit. For example, if you write down a string of equations, each successive equation should be accompanied by some justification such as “by the additivity axiom” or “because the outcomes are equally likely” or “by the distributive property of set operations.” You do not need to justify algebra.

Problem 1: Time to detection [5 pts each]

A target sits in a known cell; each sonar pass independently detects it with probability $q = 1 - \alpha$, where α is the overlook probability. Let X be the number of passes until (and including) the first detection.

- (a) Identify the distribution of X ; compute $\mathbb{E}[X]$ and $\mathbb{P}(X > k)$.
- (b) Prove the tail-sum formula

$$\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{P}(X \geq k)$$

for any nonnegative integer-valued random variable X , and use it to re-derive $\mathbb{E}[X]$ for the distribution of part (a).

- (c) Two autonomous vehicles search the same cell independently: on each pass, each vehicle independently detects the target with probability q . Let T be the time of first detection (by either vehicle). Find the distribution of T and $\mathbb{E}[T]$. Then generalize to m vehicles.

Problem 2: The sweep [5 pts each]

A sweep covers $n = 40$ cells; each cell independently contains a mine with probability 0.1, and the sensor overlooks a present mine with probability 0.25. Let X be the number of mines detected in one sweep.

- (a) Using indicator variables, find $\mathbb{E}[X]$ and $\text{Var}(X)$.
- (b) A second, independent sweep is performed over the same field. Let Y be the number of mines detected *for the first time* in sweep 2. Find $\mathbb{E}[Y]$.

Problem 3: The lying oracle [5 pts each]

A number θ is equally likely to be any element of $\{1, \dots, 16\}$. You ask “is $\theta > 8$?” and the answer is correct with probability 0.8 and wrong with probability 0.2. The answer comes back “yes.”

- (a) Compute the posterior PMF of θ (it takes two values; find both and check normalization).
- (b) You ask a second question and may pick any threshold. Show that the threshold splitting the *posterior mass* most evenly is *not* the midpoint 8, and find it.
- (c) Now suppose you keep asking threshold questions “is $\theta > \tau$?”, choosing each threshold however you like, and that given θ , each answer is correct with probability 0.8 independently of all earlier answers. Show that after k answers,

$$\mathbb{P}(\theta = j \mid \text{answers}) \propto 4^{c_j},$$

where c_j is the number of the k answers that are consistent with $\theta = j$. Then suppose your second question, “is $\theta > 11$?”, comes back “no.” Write down the posterior PMF and find the threshold that splits the posterior mass most evenly for a third question.

Problem 4: Every cell at least once [5 pts each]

A search region has 10 cells. A prior map says that, if a target is present, it lies in each of five *likely* cells with probability 0.16 and in each of five *unlikely* cells with probability 0.04. Drones are dispatched one at a time; each inspects a single cell, chosen at random according to a fixed probability vector $p = (p_1, \dots, p_{10})$, independently of all other drones. Two dispatch policies are proposed: (i) *follow the map*, where p_j is the prior probability of cell j , and (ii) *uniform*, where $p_j = 1/10$ for every cell.

The region may be declared clear only after every cell has been inspected at least once. Suppose the region is in fact empty, so drones keep coming until that happens. Let U_k be the number of cells not yet inspected after k drones, and let T be the number of drones needed until every cell has been inspected.

- (a) Using indicator variables, find $\mathbb{E}[U_k]$ for a general p . Write it out for both policies and evaluate both at $k = 30$.
- (b) Show that $\mathbb{1}\{U_k \geq 1\} \leq U_k$ for every outcome, and conclude that $\mathbb{P}(T > k) \leq \mathbb{E}[U_k]$. Use this bound to find, for each policy, the smallest number of drones k for which it guarantees $\mathbb{P}(T \leq k) \geq 0.99$. Which cells control the answer for the follow-the-map policy, and why?
- (c) For the uniform policy, write $T = G_1 + G_2 + \cdots + G_{10}$, where G_i counts the drones dispatched while exactly $i - 1$ distinct cells have been inspected (so $G_1 = 1$). Identify the distribution of each G_i and compute $\mathbb{E}[T]$. *Hint:* condition on everything that happened up to the moment the $(i - 1)$ st distinct cell is inspected. Explain why, for every $m \geq 0$,

$$\mathbb{P}(G_i > m \mid \text{that history}) = \left(\frac{i-1}{10}\right)^m,$$

and what it tells you that the right side does not depend on the history. Then explain in one or two sentences why the same decomposition does not give $\mathbb{E}[T]$ directly for the follow-the-map policy, even though the indicator argument of part (a) worked for both policies.

Problem 5: Race of the strategies [10 pts]

On the 10×10 grid from Homework 1 Problem 5, compare two search strategies over 500 random trials: (i) a fixed grid sweep, and (ii) greedy posterior maximization. Report the mean and median number of looks to detection.

Before running: predict which strategy wins and whether the gap grows or shrinks as α increases from 0.1 to 0.5.

Turn in: a plot of mean detection time versus α for both strategies, your prediction versus the outcome, and your code.

Problem 6: Audit the TA-bot [5 pts]

The assistant was asked: “A cell either contains a target (probability 0.5) or is empty. Each pass over the cell detects a present target with probability $q = 0.35$, independently across passes. After 6 failed passes: what should we believe, and how many more passes should we expect to need if we keep searching?” It answered:

TA-bot writes: Let X be the number of passes to detection when the target is present; X is geometric with success probability 0.35, so $\mathbb{P}(X > m + k \mid X > m) = \mathbb{P}(X > k)$: the geometric distribution is memoryless. Because the search process is memoryless, the six failures tell us nothing. Our belief that the target is in the cell is therefore still 0.5, and the expected number of additional passes needed is unchanged at $1/0.35 \approx 2.86$.

This solution contains **exactly one error**, and note that not every clause is wrong.

- (a) Identify the first line that goes wrong.
- (b) Name the violated principle.
- (c) Give the corrected belief.
- (d) One sentence: why is this a mistake a language model plausibly makes?