Name: Zayd Hammoudeh

Assignment: CMPS218 Take Home Final Exam

Other Student Discussions: Below is a summary of my discussions with other students for this exam.

- *Problem #1*: This problem was discussed with Konstantinos Zampetakis, Will Bolden, and Noujan Pashanasangi.
- *Problem #5*: The Seven Scientists – This problem originally appeared in the homework assignment. At the time, I discussed it with William Bolden, Konstantinos Zampetakis, and Keller Jordan. My answer is identical to the homework other than perhaps a couple of grammar corrections.

Problem Assessments: All problems are complete to the best of my understanding. I am not fully satisfied with my answer to the first problem on the twelve balls. It was quite challenging, and I am not sure how well my estimate aligns with the true number of weights required.

Chapter #4, Exercise #14

You are given 12 balls and the three-outcome balance of exercise 4.1; this time, two of the balls are odd; each odd ball may be heavy or light, and we don't know which. We want to identify the odd balls and in which direction they are odd.

Assumptions:

- The problem states, "each odd ball may be heavy or light." This language implies that all heavy odd balls have equivalent weight. Similarly, the weights of all light odd balls are equivalent.

Answer:

The hypothesis class, $\mathcal{H}$, is the set of all valid allocations of odd ball(s) in the problem. Given a three-outcome balance, the *lower bound* for the required number of weights is:

$$\Omega\left[\#\text{Weights}\right] = \lceil \log_3 |\mathcal{H}| \rceil. \tag{1}$$

The *expected* number of weights can be estimated via:

$$\mathbb{E}\left[\#\text{Weights}\right] \approx \left\lceil \frac{\log_2 |\mathcal{H}|}{H} \right\rceil, \tag{2}$$

where H is the binary entropy of the balance outcomes. Eq. (2)'s correctness is quite intuitive. Representing each member of $\mathcal{H}$ requires $\log_2 |\mathcal{H}|$ bits. Binary entropy is the *expected* information gain per weighing (in bits). The ratio of the two in turn yields the *expected number of weights*.

The quality of the estimate Eq. (2) provides is built on two assumptions. First, since entropy H is used as a normalizer, the equation assumes that H remains constant in each weighing. Next, if the probability distribution associated with the balance outcomes is far from uniform, then there will be a large divergence between this expected quantity and the upper bound (which is more relevant for Mackay's question).

To verify the Eq. (2)'s applicability, consider the case of a single odd ball. The cardinality of $\mathcal{H}$ is $2 * \binom{12}{1} = 24$. Each outcome of the balance partitions $\mathcal{H}$ into subsets of roughly equal probability. This is illustrated in Table 1 which shows the correspondence between the odd ball's placement and the balance outcome. Using Eq. (2), we estimate the expected number of weights required is approximately

$$\left\lceil \frac{\log_2 24}{1.585} \right\rceil \approx \left\lceil \frac{4.584}{1.585} \right\rceil \approx \lceil 2.893 \rceil = 3.$$

This estimate equals the actual number of weights as shown in Figure 4.2 of Mackay's text.

Table 1: Three-outcome balance position for a single odd ball that is light or heavy

Case #	Odd Ball		Balance
	Weight	Location	Position
1	Light	Left	¯|_
2	Heavy	Right	¯|_
3	Light	Right	_|¯
4	Heavy	Left	_|¯
5	Light	Neither	–|–
6	Heavy	Neither	–|–

(a) ***Estimate* how many weights are required by the optimal strategy. And what if there are three odd balls?**

When there are two odd balls, $|\mathcal{H}|$ is $2 \cdot 2 \cdot \binom{12}{2} = 264$. Table 2 shows the possible balance locations for two balls. Since this part of the problem provides no information about the relative ball weights, it is not possible to know *a priori* the balance's behavior when there are mismatched balls on the same side of the balance.

The column labeled "Probability" in Table 2 denotes the likelihood of the corresponding balance configuration. P_1 and P_2 equal $\frac{3}{154}$ and $\frac{4}{121}$, respectively.[1] Table 3 lists the grouped probability of each balance outcome. Using this table to calculate binary entropy is problematic as the "Unknown" cases do not fit neatly into any three of the balance outcomes. For simplicity, we spread the probability for the "Unknown" cases equally across the three outcomes. Therefore, this *approximated entropy* equals 1.552. Using Eq. (2), we estimate the expected number of weights required for two odd balls is:

$$\left\lceil \frac{\log_2 264}{H_{approx}} \right\rceil \approx \left\lceil \frac{8.044}{1.552} \right\rceil \approx \lceil 5.182 \rceil = \boxed{6}.$$

When there are three odd balls, $|\mathcal{H}|$ equals $2 \cdot 2 \cdot 2 \cdot \binom{12}{3} = 1,760$. Estimating the entropy is made more challenging than two odd balls as the probability of the "Unknown" configurations rises. To simplify the calculation, we reuse the approximated entropy from the two ball case. This results in an estimate of

$$\left\lceil \frac{\log_2 1760}{H_{approx}} \right\rceil \approx \left\lceil \frac{10.78}{1.552} \right\rceil \approx \lceil 6.945 \rceil = \boxed{7}$$

for the expected number of weights.

To estimate the worst case number of weights, Eq. (2) changes slightly as shown below.

$$O\left[\#\text{Weights}\right] = \left\lceil \frac{\log_2 |\mathcal{H}|}{-\log_2 \left(\max\{\Pr(\text{Balance})\}\right)} \right\rceil \tag{3}$$

[1] These values are based on four balls each in the left and right balance as well as four off to the side.

Table 2: Three-outcome balance position for two odd balls that are light and/or heavy Ball

Case #	Odd Ball #1		Odd Ball #2		Balance Position	Pr
	Weight	Location	Weight	Location		
1	Light	Left	Light	Left	$\overline{\ \ }\vert\underline{\ \ }$	P_1
2	Light	Left	Light	Neither	$\overline{\ \ }\vert\underline{\ \ }$	P_2
3	Light	Left	Heavy	Neither	$\overline{\ \ }\vert\underline{\ \ }$	P_2
4	Light	Neither	Light	Left	$\overline{\ \ }\vert\underline{\ \ }$	P_2
5	Heavy	Neither	Light	Left	$\overline{\ \ }\vert\underline{\ \ }$	P_2
6	Heavy	Right	Heavy	Right	$\overline{\ \ }\vert\underline{\ \ }$	P_1
7	Heavy	Right	Light	Neither	$\overline{\ \ }\vert\underline{\ \ }$	P_2
8	Heavy	Right	Heavy	Neither	$\overline{\ \ }\vert\underline{\ \ }$	P_2
9	Light	Neither	Heavy	Right	$\overline{\ \ }\vert\underline{\ \ }$	P_2
10	Heavy	Neither	Heavy	Right	$\overline{\ \ }\vert\underline{\ \ }$	P_2
11	Light	Right	Heavy	Left	$\overline{\ \ }\vert\underline{\ \ }$	P_2
12	Heavy	Left	Light	Left	$\overline{\ \ }\vert\underline{\ \ }$	P_1
13	Light	Right	Light	Right	$\underline{\ \ }\vert\overline{\ \ }$	P_1
14	Light	Right	Light	Neither	$\underline{\ \ }\vert\overline{\ \ }$	P_2
15	Light	Right	Heavy	Neither	$\underline{\ \ }\vert\overline{\ \ }$	P_2
16	Light	Neither	Light	Right	$\underline{\ \ }\vert\overline{\ \ }$	P_2
17	Heavy	Neither	Light	Right	$\underline{\ \ }\vert\overline{\ \ }$	P_2
18	Heavy	Left	Heavy	Left	$\underline{\ \ }\vert\overline{\ \ }$	P_1
19	Heavy	Left	Light	Neither	$\underline{\ \ }\vert\overline{\ \ }$	P_2
20	Heavy	Left	Heavy	Neither	$\underline{\ \ }\vert\overline{\ \ }$	P_2
21	Light	Neither	Heavy	Left	$\underline{\ \ }\vert\overline{\ \ }$	P_2
22	Heavy	Neither	Heavy	Left	$\underline{\ \ }\vert\overline{\ \ }$	P_2
23	Light	Left	Heavy	Right	$\underline{\ \ }\vert\overline{\ \ }$	P_2
34	Heavy	Right	Light	Right	$\underline{\ \ }\vert\overline{\ \ }$	P_1
25	Light	Right	Light	Left	$-\vert-$	P_2
26	Light	Left	Light	Right	$-\vert-$	P_2
27	Heavy	Right	Heavy	Left	$-\vert-$	P_2
28	Heavy	Left	Heavy	Right	$-\vert-$	P_2
29	Light	Neither	Light	Neither	$-\vert-$	P_1
30	Light	Neither	Heavy	Neither	$-\vert-$	P_1
31	Heavy	Neither	Light	Neither	$-\vert-$	P_1
32	Heavy	Neither	Heavy	Neither	$-\vert-$	P_1
33	Light	Left	Heavy	Left	**Unknown**	P_1
34	Heavy	Left	Right	Left	**Unknown**	P_1
35	Light	Right	Heavy	Right	**Unknown**	P_1
36	Heavy	Right	Light	Right	**Unknown**	P_1

Table 3: Probability partition by balance outcome for two odd balls of unknown relative weights

Balance Position	$\overline{\ \ }\vert\underline{\ \ }$	$\underline{\ \ }\vert\overline{\ \ }$	$-\vert-$	Unknown
Probability	$\frac{603}{1694} \approx 0.356$	$\frac{603}{1694} \approx 0.356$	$\frac{178}{847} \approx 0.210$	$\frac{12}{154} \approx 0.078$

Table 4: Updated probability partition with known ball weights in part (b)

Balance Position	$^{-}\vert_{\_}$	$_{\_}\vert^{-}$	$-\vert-$
Probability	$\frac{669}{1694} \approx 0.395$	$\frac{669}{1694} \approx 0.395$	$\frac{178}{847} \approx 0.210$

Using Table 3, the Pr(Balance) equals 0.356. Accounting for the equal spreading of the "Unknown" probability, we will use probability 0.382 in the calculations below. Therefore, the worst case number of weighings for two balls is estimated as:

$$O\left[\#\text{Weights Two Balls}\right] \approx \left\lceil \frac{\log_2 264}{-\log_2 0.382} \right\rceil \approx \lceil 5.792 \rceil = \boxed{6}.$$

Similarly, for three balls, the upper bound is:

$$O\left[\#\text{Weights Three Balls}\right] \approx \left\lceil \frac{\log_2 1760}{-\log_2 0.382} \right\rceil \approx \lceil 7.766 \rceil = \boxed{8}.$$

Conclusion:

For two odd balls, both the expected and upper bound calculations estimated that 6 weights are required. This provides very high confidence in the estimation. When there are three odd balls, the upper bound of 8 weights appears to be a more appropriate estimation.

(b) How do your answers change if it is known that all the regular balls weigh 100g, that light balls weight 99g, and heavy ones weigh 110g?

In part (a), uncertainty about the balance's behavior in some cases necessitated that we approximate the binary entropy. Now that the relative ball weights are known, Cases #33-34 and Cases #35-36 in Table 2 are assigned to "$_{\_}\vert^{-}$" and "$^{-}\vert_{\_}$" respectively. The updated probability breakdown for each balance outcome is shown in Table 4. This corresponds to a binary entropy of 1.531. Note that even though we have more information, the binary entropy went down. That is because when approximating the entropy in part (a), we spread the probability corresponding to the "Unknown" case evenly across the three balance outcomes. This lead to a more balanced probability mass function (pmf), which in turn led to a higher entropy.

Using the updated entropy, the expected number of weights for two balls remains 6, but the worst case was $\lceil 6.003 \rceil = 7$. Given that the worst case is only marginally greater than six before applying the ceiling function, we hypothesize that six measurements is the better estimate.

For three balls, the expected number of weights increased to $\lceil 7.042 \rceil = 8$ and the worst case estimate was $\lceil 8.045 \rceil = 9$. Both the expected and worst case estimates are only slightly more than their previous estimate. We estimate that splitting the difference and estimating 8 weights is the best estimate for three balls.

Chapter #5, Exercise #24
Write a short essay describing how to play the game of twenty questions optimally. [In twenty questions, one player thinks of an object, and the other player has to guess the object using as few binary question as possible, preferably fewer than twenty.]

The naïve approach to answering this question is as follows:

> *There is a finite set of candidates, C, from which the first player will select the object. With each question, reduce the size of this set of possible objects by half.*

This strategy overlooks one key detail – each object in C does not have equal likelihood of selection. What is more, the likelihood of selection for each object varies based on the specific person doing the selection! For example, I am more likely to select my cat as the object than you are (whoever is reading this).

A Better Strategy:

For each $c \in C$, the "other" player should define a probability which represents the likelihood that the first player would select c. As in the naïve strategy, the set of possible candidates shrinks with each question. However, rather than reducing the size of the candidate pool in half each time, the question posed should partition that pool such that the probability of all candidates that remain valid if the answer is "no" has **equal probability** to the set of candidates that remain if the answer is "yes." This approach maximizes the expected information gain of each answer.

Once one remaining possible object, c, has greater probability than all other remaining objects combined, the "other" player should ask "Is your object c?"

Chapter #15, Exercise #5

In a magic trick, there are three participants: the magician, an assistant, and a volunteer. The assistant, who claims to have paranormal abilities, is in a soundproof room. The magician gives the volunteer six blank cards, five white and one blue. The volunteer writes a different integer from 1 to 100 on each card, as the magician is watching. The volunteer keeps the blue card. The magician arranges the white cards in some order and passes them to the assistant. The assistant then announces the number on the blue card.

How does this trick work?

For the sake of making this problem meaningful, I am going to *assume* the assistant does not actually have paranormal abilities.

Consider a set of five distinct, arbitrary integers. Assign these five integers to the letters a, b, c, d, and e satisfying the relationship $a < b < c < d < e$. At this point, the original numbers can be forgotten and only the five letters considered since they will be all that matters. It is elementary to see that there are ${}_5P_5 = 5! = 120$ permutations of five unique letters. Assign to each permutation a unique number from 1 to 120. This represents a uniquely decodable encoding of any number between 1 and 120.

In this trick, the five unique integers written on the white cards are bijectively mapped to the letters a through e as described above. The magician saw the number written on the blue card and encodes that number based on how he orders (permutes) the five white cards (letters) before passing them to the assistant. The assistant then decodes this permutation to get the volunteer's number on the blue card completing the trick – no paranormal abilities needed.

Chapter #15, Exercise #6
How does *this* trick work?

'Here's an ordinary pack of cards shuffled into random order. Please choose five cards from the pack, any you wish. Don't let me see their faces. No, don't give them to me: pass them to my assistant Esmerelda. She can look at them.

'Now, Esmerelda, show me four of the cards. Hmm...nine of spades, six of clubs, four of hearts, ten of diamonds. The hidden card must be the queen of spades.

This trick can be performed as described above for a pack of 52 cards. Use information theory to give an upper bound on the number of cards for which the trick can be performed.

Assumption:

To make this problem more meaningful, I am going to assume that when Esmerelda "shows" the four cards to the magician, she is not using any sleight of hand trickery to relay additional information to the magician (e.g., orient the cards in her hand in a certain way, show the face or backside first, etc.). Essentially, I consider that the magician is only provided numbers from a black box.

Answer:

Consider a collection of n unique objects. It is trivial to bijectively map these objects to the integers $\{1, \ldots, n\}$. If five integers from the set are selected of which four are revealed, then the previous problem's trick allows us to communicate the fifth object as long as n is less than or equal to $4! + 4 = 28$. A standard deck contains $n = 52$ cards so this is clearly insufficient to uniquely encode the hidden, fifth card. Therefore, we need a more creative encoding....

The key to this problem is that Esmerelda *chose* which of the five cards remained hidden. If she chose that card randomly, we are back to 24 permutations. However, Esmerelda was more cunning and chose a *specific* card to remain hidden. A scheme could be developed that uses this selection process to increase the upper bound on the number of encodable objects fivefold i.e., once for each card she could have chosen. When we combine that knowledge with the technique from the last question (exercise 15.5), we see that the maximum number of objects that can be uniquely encoded is $5 \cdot 4! = 120$. Therefore, combining the maximum size of the encodable objects pool with the four cards Esmerelda revealed, we see the upper bound for the maximum deck size is $4 + 120 = \boxed{124}$.

The problem does not require that we prove the tightness of this upper bound. However, after completing the problem, I did research online and found that this is in fact the true bound. What is more, the technique described above is not how this trick (known as the "Fitch Cheney Card Trick") is traditionally performed; that approach to the trick has a maximum deck size of 52.

Chapter #22, Exercise #5

15 The seven scientists. N datapoints $\{x_n\}$ are drawn from N distributions, all of which are Gaussian with a common mean μ but with different unknown standard deviations σ_n. What are the maximum likelihood parameter μ,$\{\sigma_n\}$ given the data? For example, seven scientists (A, B, C, D, E, F, G) with wildly-differing experimental skills to measure μ. You expect some of them to do accurate work (i.e., to have small σ_n), and some of them to turn in wildly inaccurate results (i.e., to have enormous σ_n). Table 5 shows their seven results. What is the μ, and how reliable is each scientist?

Scientist	x_n
A	−27.020
B	3.570
C	8.191
D	9.898
E	9.603
F	9.945
G	10.056

Table 5: Seven measurements $\{x_n\}$ of a parameter μ by seven scientists each having his own noise-level σ_n.

I hope that you agree that, intuitively, it looks pretty certain that A and B are both inept measurers, that D-G are better, and that the true value μ is somewhere close to 10. But what does maximizing the likelihood tell you?

Given μ, the probability that observer i with measurement standard deviation, σ_n, measures any single value, x_n, is:

$$\Pr\left(x_n|\mu,\sigma_n\right) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-(x_n-\mu)^2}{2\sigma_n^2}\right).$$

If there are N (independent) observers, then the likelihood, L, of the measurements $\{x_n\}$ where $n \in \{1,\dots,N\}$ is:

$$\begin{aligned} L &= \Pr\left(\{x_n\}|\mu,\{\sigma_n\}\right) \\ &= \prod_{n=1}^{N} \Pr\left(x_n|\mu,\sigma_n\right) \\ &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-(x_n-\mu)^2}{2\sigma_n^2}\right). \end{aligned}$$

Taking the natural logarithm of this function yields:

$$\ln(L) = \sum_{n=1}^{N} \left(-\frac{1}{2}\ln\left(2\pi\sigma_n^2\right) - \frac{(x_n-\mu)^2}{2\sigma_n^2}\right).$$

Since the natural logarithm function is strictly increasing, we can take the derivative of $\ln(L)$ with respect to μ to find the maximum likelihood for μ. This yields:

$$
\begin{aligned}
0 &= \sum_{n=1}^{N} \frac{(x_n - \mu)}{\sigma_n^2} \\
\mu &= \frac{1}{\sum_{n=1}^{N} \sigma_n^{-2}} \sum_{n=1}^{N} \frac{x_n}{\sigma_n^2}.
\end{aligned}
$$

For each σ_n, the maximum likelihood estimate is found by taking the partial derivative with respect to σ_n yielding:

$$
\begin{aligned}
0 &= -\frac{2\pi\sigma_n}{4\pi\sigma_n^2} + \frac{(x_i - \mu)^2}{\sigma_n^3} \\
\sigma_n &= \sqrt{(x_n - \mu)^2} = |x_n - \mu|.
\end{aligned}
$$

To find the maximum likelihood, $L_{\max}$, substitute the definition for σ_n found above. This yields

$$
\begin{aligned}
L_{\max} &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi(x_n - \mu)^2}} \exp\left(\frac{-(x_n - \mu)^2}{2(x_n - \mu)^2}\right) \\
&= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi(x_n - \mu)^2}} \exp\left(-\frac{1}{2}\right).
\end{aligned}
$$

Therefore, the likelihood is maximized when $\mu = x_n$. For this problem, x_n is any value in Table 5 meaning the maximum likelihood is when any of the seven scientists are perfectly correct, i.e., always report the true μ with no variation. This result is disquieting as it means the likelihood is maximized if μ is -27.020 or 3.570. As noted in the question itself, the visceral feeling is that the true μ is around ten give or take. Therefore, maximizing the likelihood may not always yield the best result.