Acknowledgement: The TA graciously thanks Rafael Stern for providing most of the solutions.

1.1 Problem 1

1.1.1 1.1

Weighing 6 balls vs 6 balls yields only two possible outcomes: left heavier or right heavier. Hence, any weighing scheme that at any point weighs 6 vs 6 can yields, in 3 weighings, at most $2 \times 3 \times 3 = 18$ possible outcomes. However, to distinguish the special ball and determine whether it is heavier or lighter, we need to our method to differentiate between 24 possible cases. $18 < 24$, and hence, we need at least 4 weighings. Since the optimal method uses 3 weighings, this is suboptimal.

1.1.2 1.2

We can use similar reasoning as 1.1.

Suppose that the outcome of the first weighing is that the two sets of 3 balls are equally heavy, then the odd ball must be in the remaining 6 balls. We must then be able to differentiate between 12 cases. Therefore, we need at least 3 additional weighings to find the odd ball in the remaining 6 balls.

We can then conclude that any method that weighs 3 balls versus 3 balls in its first weighing needs at least 4 weighings.

1.1.3 1.3

See figure 1.1.3

1.2 Problem 2

1.2.1 2.1

To prove that $H(X|Y) \leq H(X)$, recall that $H(X) = H(X|Y) + I(X;Y)$. Hence, it is enough to show that $I(X;Y) \geq 0$. Observe that $I(X;Y) = D_{KL}(p(x,y); p(x)p(y))$. Thus, by Gibb’s Inequality, $I(X;Y) \geq 0$ and the desired result is established.

Next, say that a variable has distribution $S(p)$ if $P(X = 1) = p$ and $P(X = -1) = 1 - p$. Define $X \sim S(0.5)$ and $Z \sim S(0.5)$ independent random variables. Consider $Y = XZ$. Observe that $Y$ is independent of $X$ and, thus, $I(X;Y) = 0$. Observe that, given no value of $Z$, $X$ and $Y$ are independent, that is, $I(X;Y|Z = z) > 0$. Thus, $I(X;Y|Z) > 0 = I(X;Y)$.
1.2.2 2.2

Let’s prove this result by induction. The case $n = 2$ was shown in class, hence it is only necessary to prove the inductive step. Let’s assume that the equality works for $n$ and show that it works for $n + 1$.

$$H(X_1, \ldots, X_{n+1}) = -\sum_{x_1, \ldots, x_{n+1}} p(x_1, \ldots, x_{n+1}) \log(p(x_1, \ldots, x_{n+1})) =$$

$$= -\sum_{x_1, \ldots, x_{n+1}} p(x_1, \ldots, x_{n+1}) \log(p(x_{n+1})) - \sum_{x_1, \ldots, x_n} p(x_1, \ldots, x_n) \log(p(x_1, \ldots, x_n| x_{n+1})) =$$

$$= -\sum_{x_{n+1}} p(x_{n+1}) \log(p(x_{n+1})) - \sum_{x_{n+1}} p(x_{n+1}) \sum_{x_1, \ldots, x_n} p(x_1, \ldots, x_n| x_{n+1}) \log(p(x_1, \ldots, x_n| x_{n+1})) =$$

$$= H(X_{n+1}) + \sum_{x_{n+1}} p(x_{n+1}) H(p(x_1, \ldots, x_n| x_{n+1})) =$$

Calling $p(x_{a_1}, \ldots, x_{a_n}| x_{b_1}, \ldots, x_{b_n}, x_{n+1}) = q_{x_{n+1}}(x_{a_1}, \ldots, x_{a_n}| x_{b_1}, \ldots, x_{b_n})$

$$= H(X_{n+1}) + \sum_{x_{n+1}} p(x_{n+1}) H(q_{x_{n+1}}(x_1, \ldots, x_n)) =$$

By the induction hypothesis:

$$\sum_{x_{n+1}} p(x_{n+1}) H(q_{x_{n+1}}(x_1, \ldots, x_n)) = \sum_{x_{n+1}} p(x_{n+1}) \sum_{i} H(q_{x_{n+1}}(x_{n-i}| x_n, \ldots, x_{n-i+1})) =$$

$$= \sum_{i\geq 1} H(X_{n+1-i}|X_{n+1}, \ldots, X_{n+2-i})$$

Substituting this expression in the main equality:

$$H(X_1, \ldots, X_{n+1}) = \sum_{i\geq 0} H(X_{n+1-i}|X_{n+1}, \ldots, X_{n+2-i})$$

Which completes the proof. Equality holds when $X_1, \ldots, X_n$ are jointly independent, that is, any two disjoint subsets of $(X_1, \ldots, X_n)$ are independent.

1.2.3 2.3

$$I(X_1, \ldots, X_n; Y) = H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n| Y) =$$

$$= H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n, Y) - H(Y) =$$

Applying item (b) twice:
1.3 Problem 3

by Exercise 2a, the proof is complete.

1.2.4 2.4

Note: many of you did not prove the first part—that \( D(\cdot \| \cdot) \) is convex. Rafael’s solution is very slick and I recommend that you take a look if you hadn’t proved it yourself.

\[
D_{KL}(\lambda p_1 + (1 - \lambda)p_2 || \lambda q_1 + (1 - \lambda)q_2) \leq \lambda D_{KL}(p_1 || q_1) + (1 - \lambda)D_{KL}(p_2 || q_2)
\]

We wish to prove that \( H(\lambda p_1 + (1 - \lambda)p_2) \geq \lambda H(p_1) + (1 - \lambda)H(p_2) \). Take \( X \sim p_1 \) and \( Y \sim p_2 \). Consider \( \xi \sim \text{Ber}(\lambda) \) independent of \( X \) and \( Y \). Define \( Z = \xi X + (1 - \xi)Y \). Observe that \( Z \sim \lambda p_1 + (1 - \lambda)p_2 \). Hence the left hand side of the inequality equals \( H(Z) \). Observe that the right hand side equals \( H(Z|B) \) and, thus, by Exercise 2a, the proof is complete.

1.3 Problem 3

1.3.1 3.1

Using the data processing inequality, we know that \( I(U; V) \leq I(X; V) \). Hence, we only need to prove that \( I(X; V) \leq I(X; Y) \). By symmetry of mutual information, this is equivalent to show that \( I(V, X) \leq I(Y; X) \). If we could invert the graph as in the hint, we could establish this inequality by using the data processing inequality once again. Hence, it only remains to prove that \( X \to Y \to Z \) implies \( Z \to Y \to X \).

Assume \( X \to Y \to Z \). Hence, \( p(x, y, z) = p(x)p(y|x)p(z|y) \) Using Bayes Rule, conclude that: \( p(x, y, z) = p(x|y)p(y)p(z|y) \). Finally, using Bayes Rule again: \( p(x, y, z) = p(x|y)p(y|z)p(z) \). Since \( x, y \) and \( z \) were arbitrary, we have established that \( Z \to Y \to X \), which completes the proof.

1.3.2 3.2

\[
f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} I_{[\theta, \infty)}(x_i)I_{(-\infty, \theta+1]}(x_i) =
\]
\[
= \prod_{i=1}^{n} I_{[0, \infty]}(x_i) \prod_{i=1}^{n} I_{[-\infty, \theta+1]}(x_i) = I_{[0, \infty]}(x_{(1)}) I_{[-\infty, \theta+1]}(x_{(n)})
\]

Hence, by Fisher’s Factorization Theorem, \( T \) is a sufficient statistic for \( \theta \) in the statistical sense. Hence, by the result seen during Recitation 1, it is also sufficient in the information theoretic sense.
Figure 1.1: Solution to Problem 1.3

Either \{A, B\} light or \{E\} heavy

\[ \begin{cases} \text{A if light} \\ \text{B if heavy} \end{cases} \]

Either \{D, I\} light or \{G, H\} heavy

\[ \begin{cases} \text{D if light} \\ \text{G if heavy} \end{cases} \]

Either \{E\} light or \{C\} heavy

\[ \begin{cases} \text{E if light} \\ \text{C if heavy} \end{cases} \]

Either \{F\} heavy or \{A\} light

\[ \begin{cases} \text{F if heavy} \\ \text{A if light} \end{cases} \]

Either \{B\} light

\[ \begin{cases} \text{B if light} \end{cases} \]

Either \{J\} light or \{J, K\} heavy

\[ \begin{cases} \text{J if light} \\ \text{J, K if heavy} \end{cases} \]

Either \{L\} light or \{L\} heavy

\[ \begin{cases} \text{L if light} \\ \text{L if heavy} \end{cases} \]