ECE 563 FA25 HW2 Solutions

Problem 1. Properties of mutual information.

Solution.

(a) **No**, there exist random variables X, Y_1, and Y_2 such that $I(X;Y_1) = I(X;Y_2) = 0$ while $I(X;Y_1,Y_2) \neq 0$.
Similar to the solution to Problem 8 of HW1, one way to find such a counterexample is to consider pairwise independent but not mutually independent random variables. Let Y_1 and Y_2 to be i.i.d. Bernoulli($\frac{1}{2}$) random variables, and let X be the indicator function of the event $Y_1 = Y_2$. Then, it can be checked that X and Y_1 are independent, and so are X and Y_2. These properties imply that $I(X;Y_1) = I(X;Y_2) = 0$. At the same time, X is clearly a function of Y_1 and Y_2, and we can check that X is a Bernoulli($\frac{1}{2}$) random variable as well. Therefore, we have $I(X;Y_1,Y_2) = H(X) = 1 \neq 0$.

(b) **No**, there exist random variables X, Y_1, and Y_2 such that $I(X;Y_1) = I(X;Y_2) = 0$ while $I(Y_1;Y_2) \neq 0$.
We can simply consider X, Y_1 to be i.i.d. Bernoulli($\frac{1}{2}$) random variables and let $Y_2 = Y_1$. It then follows that $I(X;Y_1) = I(X;Y_2) = 0$ and $I(Y_1;Y_2) = H(Y_1) = 1 \neq 0$.

Problem 2. Data Processing Inequality.

Solution.

(d) We first prove Part (d). We start with the formula

$$\begin{aligned}
I(X;Z|Y) &= H(X|Y) + H(Z|Y) - H(X,Z|Y) \\
&= -\sum_{x,y} p(x,y)\log p(x|y) - \sum_{y,z} p(y,z)\log p(z|y) + \sum_{x,y,z} p(x,y,z)\log p(x,z|y) \\
&= -\sum_{x,y,z} p(x,y,z)\log p(x|y) - \sum_{x,y,z} p(x,y,z)\log p(z|y) + \sum_{x,y,z} p(x,y,z)\log p(x,z|y) && (2.1) \\
&= \sum_{x,y,z} p(x,y,z)\log \frac{p(x,z|y)}{p(x|y)p(z|y)}, && (2.2)
\end{aligned}$$

where in (2.1) we used the fact that $p(x,y) = \sum_z p(x,y,z)$ and $p(y,z) = \sum_x p(x,y,z)$. Then, by the assumption that $X \to Y \to Z$, we have for all x, y, z that $p(x,z|y) = p(x|y)p(z|y)$. Thus, (2.2) becomes

$$\begin{aligned}
I(X;Z|Y) &= \sum_{x,y,z} p(x,y,z)\log \frac{p(x|y)p(z|y)}{p(x|y)p(z|y)} \\
&= \sum_{x,y,z} p(x,y,z)\log 1 \\
&= 0.
\end{aligned}$$

(a) Note that

$$H(X,Z|Y) = H(X|Y,Z) + H(Z|Y), \quad (2.3)$$
$$H(X,Z|Y) = H(Z|X,Y) + H(X|Y). \quad (2.4)$$

Comparing (2.3) and (2.4), we can see that $H(X|Y) = H(X|Y,Z)$ if and only if

$$H(Z|Y) = H(Z|X,Y). \quad (2.5)$$

But note that from Part (d) we have $I(X;Z|Y) = H(Z|Y) - H(Z|X,Y) = 0$. Therefore (2.5) holds, which implies $H(X|Y) = H(X|Y,Z)$.

(b) From Part (a) and the fact that conditioning reduces entropy, we have

$$\begin{aligned}
H(X|Y) &= H(X|Y,Z) \\
&\leq H(X|Z).
\end{aligned}$$

(c) From Part (b), we have

$$\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&\geq H(X) - H(X|Z) \\
&= I(X;Z).
\end{aligned}$$

Problem 3. Divergence. Prove that

$$d(p||q) \geq 2(p-q)^2 \log e. \tag{3.1}$$

Solution.

We first prove (3.1) for $p, q \in (0,1)$.

Fix $p \in (0,1)$. Consider the function

$$\begin{aligned} f(q) &:= d(p||q) - 2(p-q)^2 \log e \\ &= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} - 2(p-q)^2 \log e, \end{aligned} \tag{3.2}$$

which is defined for $q \in (0,1)$. Taking the derivative of $f(q)$ in (3.2) with respective to q gives

$$\begin{aligned} f'(q) &= (\log e)(-\frac{p}{q} + \frac{1-p}{1-q} + 4(p-q)) \\ &= (\log e)\frac{(q-p)(2q-1)^2}{q(1-q)}. \end{aligned} \tag{3.3}$$

From (3.3) we can see that

- $f'(q) \leq 0$ for $q < p$, and
- $f'(q) \geq 0$ for $q > p$.

This implies

- $f(q)$ is non-increasing on the interval $(0,p)$, and
- $f(q)$ is non-decreasing on the interval $(p,1)$.

We can thus deduce that $f(q)$ attains a global minimum at $q = p$. That is, we have for any $q \in (0,1)$ that

$$\begin{aligned} f(q) &\geq f(p) \\ &= d(p||p) - 2(p-p)^2 \log e \\ &= 0, \end{aligned}$$

which proves (3.1) for $p, q \in (0,1)$.

Then consider $p \in \{0,1\}$ and $q \in (0,1)$. We will only show the derivations for the case $p = 0$, and the case $p = 1$ can be done similarly. Note that we can still construct the same function $f(q)$ as in (3.2), but now we can extend the domain of $f(q)$ to be $[0,1)$ (since $0 \log \frac{0}{0} = 0$ by convention). A similar computation to that in (3.3) shows that $f'(q) \geq 0$ for all $q \in [0,1)$, and thus $f(q)$ attains a global minimum at $q = 0$. That is, we have $f(q) \geq f(0) = d(0||0) - 2(0-0)^2 \log e = 0$ for all $q \in [0,1)$, which in particular holds for $q \in (0,1)$.

Lastly, consider $q \in \{0,1\}$. Note that in this case we have for all $p \in [0,1]$ that

$$d(p||q) = \begin{cases} \infty, & \text{if } p \neq q, \\ 0, & \text{if } p = q, \end{cases}$$

and the inequality in (3.1) easily follows.

Problem 4. [1, Problem 3.7] "AEP-like limit."

Solution.

Note that

$$\log(p(X_1, \ldots, X_n))^{\frac{1}{n}} = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i). \tag{4.1}$$

Thus we first find the limit of $\frac{1}{n}\sum_{i=1}^{n} \log p(X_i)$. Define for each $i \geq 1$ that $Y_i := \log p(X_i)$. Since $X_1, X_2, \ldots,$ are i.i.d., we know that $Y_1, Y_2, \ldots$ are i.i.d. as well. We can then use the strong law of large numbers to deduce that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{a.s.} \mathbb{E}[Y_1], \tag{4.2}$$

where $\xrightarrow{a.s.}$ means almost sure convergence. Note that $\mathbb{E}[Y_1]$ is simply

$$\begin{aligned} \mathbb{E}[Y_1] &= \mathbb{E}[\log p(X_1)] \\ &= \sum_x p(x) \log p(x) \\ &= -H(X_1). \end{aligned} \tag{4.3}$$

That is, we have from (4.2) and (4.3) that

$$\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) \xrightarrow{a.s.} -H(X_1). \tag{4.4}$$

It then follows from (4.1), (4.4), and the continuity of $t \mapsto 2^t$ that

$$\begin{aligned}(p(X_1,\ldots,X_n))^{\frac{1}{n}} &= 2^{\log(p(X_1,\ldots,X_n))^{\frac{1}{n}}} \\ &\xrightarrow{a.s.} 2^{-H(X_1)}.\end{aligned}$$

Problem 5. [1, Problem 3.10] "Random box size."

Solution.

Following a similar construction as in Problem 4, we define for each $n \geq 1$ that $Y_n := \ln X_n$, where $\ln(\cdot)$ denotes the natural log. It follows that $Y_1, Y_2, \ldots$ are i.i.d., and thus the strong law of large number implies that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{a.s.} \mathbb{E}[Y_1], \tag{5.1}$$

where

$$\begin{aligned}\mathbb{E}[Y_1] &= \mathbb{E}[\ln X_1] \\ &= \int_{x=0}^{1} \ln x dx \\ &= x\ln x - x|_{x=0}^{1} \\ &= -1.\end{aligned} \tag{5.2}$$

It follows from (5.1), (5.2), the definition of Y_n, and the continuity of $t \mapsto e^t$ that

$$\begin{aligned}V_n^{1/n} &= e^{\frac{1}{n}\sum_{i=1}^{n}\ln X_n} \\ &\xrightarrow{a.s.} e^{-1}.\end{aligned} \tag{5.3}$$

We now calculate $\mathbb{E}[V_n]^{\frac{1}{n}}$. By the independence of $X_1, X_2, \ldots$, we have

$$\begin{aligned}\mathbb{E}[V_n] &= \prod_{i=1}^{n}\mathbb{E}[X_i] \\ &= 2^{-n},\end{aligned}$$

where we used the fact that the expectation of a uniform $[0,1]$ random variable is $\frac{1}{2}$. It follows that $\mathbb{E}[V_n]^{\frac{1}{n}} = \frac{1}{2}$, which is different from the a.s. limit $V_n^{1/n} \xrightarrow{a.s.} e^{-1}$ in (5.3).

Problem 6. [1, Problem 5.2] "How many fingers has a Martian?"

Solution.

Since the codewords are uniquely decodable, by the McMillan inequality, we must have

$$D^{-1} + D^{-1} + D^{-2} + D^{-3} + D^{-2} + D^{-3} \leq 1,$$

or equivalently

$$D^3 - 2D^2 - 2D - 2 \geq 0. \tag{6.1}$$

Define $f(D) := D^3 - 2D^2 - 2D - 2$. It can be calculated that $f(1) = -5 < 0$, $f(2) = -6 < 0$, and $f(3) = 1 \geq 0$. At the same time, we have $f'(D) = 3D^2 - 4D - 2 > 0$ for $D \geq 3$. Therefore, the alphabet size D (which is necessarily a positive integer) satisfies (6.1) if and only if $D \geq 3$. That is, the McMillan inequality is satisfied if and only if $D \geq 3$, and thus a good lower bound on D is 3.

The preparers of these solutions do not find how the alphabet size of uniquely decodable codes is related to the number of fingers a Martian (or any species) has. We human have ten fingers, but we are using binary uniquely decodable codes everywhere.

Problem 7. Depth constraint Huffman codes.

Solution.

The paper "Near-Optimal Depth-Constrained Codes" by Gupta, Prabhakar, and Boyd [2] addresses the problem of constructing prefix codes with a maximum depth constraint L, providing efficient algorithms with provable performance guarantees.

Main Results:

1) **Theorem 1 (Huffman codes):** A depth-constrained Huffman code can be constructed in $O(n \log n)$ time and $O(n)$ space, with average codeword length within 1 bit of the optimal depth-constrained code.
2) **Theorem 2 (Alphabetic codes):** A depth-constrained alphabetic code can be constructed in $O(n \log n)$ time and $O(n)$ space, with average codeword length within 2 bits of the optimal depth-constrained alphabetic code.

Approach:

The key innovation is recasting the depth-constrained coding problem as a convex optimization problem:

1) **Probability transformation:** Given probabilities $\{p_i\}$, find transformed probabilities $\{q_i^*\}$ satisfying:
   - $\sum_i q_i = 1$
   - $q_i \geq Q = 2^{-L}$ for all i (ensures depth $\leq L$)
   - Minimize the relative entropy: $D(p||q) = \sum_i p_i \log(p_i/q_i)$
2) **Lagrange multiplier solution:** Using Lagrange multipliers, the optimal solution is:
$$q_i^* = \max(p_i/\mu^*, Q)$$
   where μ^* is found via binary search to satisfy $\sum_i q_i^* = 1$.
3) **Codeword construction:**
   - For Huffman codes: $l_i^* = \lceil -\log q_i^* \rceil$
   - For alphabetic codes: Modified lengths satisfying Yeung's characteristic inequality

Intuition:

1) **Why transform probabilities?** If $p_{\min} < 2^{-L}$, the natural code would have depth $> L$. Transforming to $q_i \geq 2^{-L}$ ensures all codewords have length $\leq L$.
2) **Why minimize relative entropy?** The average codeword length satisfies:
$$\sum_i p_i l_i^* \leq \sum_i p_i \log(1/q_i) + 1 = D(p||q) + H(p) + 1$$
   Thus minimizing $D(p||q)$ minimizes an upper bound on the average length.
3) **Geometric interpretation:** The solution $q_i^* = \max(p_i/\mu^*, Q)$ scales down probabilities uniformly but clips them at the minimum threshold $Q = 2^{-L}$.
4) **Convex optimization advantage:** The problem has a unique global optimum that can be found efficiently, avoiding the complexity of exact algorithms like Package-Merge.
5) **Trade-off:** Sacrificing 1-2 bits of optimality yields a simple, fast algorithm with no dependence on letter probabilities in the complexity.

References

[1] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[2] P. Gupta, B. Prabhakar, and S. Boyd, "Near-optimal depth-constrained codes," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3294–3298, 2004.