CS3236: Solutions to Tutorial 3
(Block Source Coding)

1. **[Typical Set Calculations]**

   (a) Suppose a discrete memoryless source (DMS) emits h (heads) and t (tails) with probability 1/2 each, i.e., $P_X(h) = P_X(t) = \frac{1}{2}$. For $\epsilon = 0.01$ and $n = 5$, what is the typical set $\mathcal{T}_n(\epsilon)$?
   (Hint: This part can be answered in one line)
   Solution. *In this case, $H(X) = 1$. All source sequences are equally likely, each with probability $2^{-5} = 2^{-nH(X)}$. Hence, all sequences satisfy the condition for being typical,*

   $$2^{-n(H(X)+\epsilon)} \le P_{\mathbf{X}}(\mathbf{x}) \le 2^{-n(H(X)-\epsilon)}$$

   for any $\epsilon > 0$. Hence, all 32 sequences are typical.

   (b) Repeat part (a) with $P_X(h) = 0.2$, $P_X(t) = 0.8$, $n = 5$, and $\epsilon = 0.0001$.
   (Hint: Try taking logs in the definition of typicality, applying the definition of entropy, and simplifying as much as possible.)
   Solution. *Consider a sequence with m heads and $n-m$ tails. Then, the probability of occurrence of this sequence is $p^m(1-p)^{n-m}$, where $p = P_X(h)$. For such a sequence to be typical*

   $$2^{-n(H(X)+\epsilon)} \le p^m(1-p)^{n-m} \le 2^{-n(H(X)-\epsilon)}.$$

   To simplify the notation, let's write this condition as "$p^m(1-p)^{n-m} = 2^{-n(H(X)+\alpha)}$ for some $\alpha \in [-\epsilon, \epsilon]$". Then we can simplify as follows by taking the log:

   $$\begin{aligned} & p^m(1-p)^{n-m} = 2^{-n(H(X)+\alpha)} \\ \iff\ & m\log_2 p + (n-m)\log_2(1-p) = -nH(X) - n\alpha \\ \iff\ & m\log_2 p + (n-m)\log_2(1-p) = np\log_2 p + n(1-p)\log_2(1-p) - n\alpha, \end{aligned}$$

   where the last line is by the definition of entropy. Canceling an $n\log_2 p$ term from both sides, and applying $\log_2 p - \log_2(1-p) = \log_2 \frac{p}{1-p}$, we obtain the condition

   $$m\log_2 \frac{p}{1-p} = -n\alpha + np\log_2 \frac{p}{1-p}, \tag{1}$$

   which (recalling $\alpha \in [\epsilon, \epsilon]$) translates to

   $$\left| \left(\frac{m}{n} - p\right) \log_2 \frac{1-p}{p} \right| \le \epsilon.$$

   Plugging in the values $p = 0.2$ and $n = 5$, and noting $\log_2 \frac{0.8}{0.2} = 2$, we get

   $$\left| \frac{m}{5} - \frac{1}{5} \right| \le \frac{\epsilon}{2}.$$

   Since $m = \{0, \dots, 5\}$, this condition will be satisfied for the given ϵ only for $m = 1$ i.e. when there is one H in the sequence. Thus,

   $$\mathcal{T}_n(\epsilon) = \{(HTTTT), (THTTT), (TTHTT), (TTTHT), (TTTTH)\}.$$

2. **[Strong Typicality]**

Consider a source distribution P_X such that $P_X(x) > 0$ for all $x \in \mathcal{X}$, where $\mathcal{X}$ is a finite alphabet. The *strongly typical* set is defined as

$$\mathcal{T}_n(\epsilon) = \Big\{ \mathbf{x} \in \mathcal{X}^n \,:\, nP_X(x)(1-\epsilon) \le n_x(\mathbf{x}) \le nP_X(x)(1+\epsilon), \quad \forall x \in \mathcal{X} \Big\},$$

where $n_x(\mathbf{x}) = \sum_{i=1}^n \mathbf{1}\{x_i = x\}$ is the number of times x occurs in the sequence $\mathbf{x} = (x_1, \dots, x_n)$. This is a bit easier to interpret than the definition of typicality from the lecture: It just states that the observed proportion of occurrences of each symbol is roughly equal to the probability of that symbol.

(a) Show that for $\mathbf{X} = (X_1, \dots, X_n)$ distributed i.i.d. on P_X, it holds that $\mathbb{P}[\mathbf{X} \in \mathcal{T}_n(\epsilon)] \to 1$ as $n \to \infty$ for fixed $\epsilon > 0$.

Solution. *Notice that $n_x(\mathbf{X})$ is a sum of i.i.d. random variables $Z_i = \mathbf{1}\{X_i = x\}$ with $\mathbb{E}[Z_i] = P_X(x)$. Therefore, by the law of large numbers, the probability that $\frac{n_x(\mathbf{x})}{n}$ deviates from $P_X(x)$ by more than $\epsilon P_X(x) > 0$ tends to zero as $n \to \infty$. Using the union bound and the fact that $\mathcal{X}$ is a finite set, we have*

$$\begin{aligned}
\mathbb{P}[\mathbf{X} \notin \mathcal{T}_n(\epsilon)] &= \mathbb{P}\Bigg[\bigcup_{x \in \mathcal{X}} \Big\{ \text{the event } nP_X(x)(1-\epsilon) \le n_x(\mathbf{X}) \le nP_X(x)(1+\epsilon) \text{ does not hold} \Big\} \Bigg] \\
&\le \sum_{x \in \mathcal{X}} \mathbb{P}\big[\text{the event } nP_X(x)(1-\epsilon) \le n_x(\mathbf{X}) \le nP_X(x)(1+\epsilon) \text{ does not hold} \big] \\
&\to 0,
\end{aligned}$$

since the sum of a fixed number of terms that tend to zero also tends to zero.

(b) Show that for any non-negative valued function $a(x)$, and any sequence $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{T}_n(\epsilon)$, it holds that

$$\mathbb{E}[a(X)](1-\epsilon) \le \frac{1}{n} \sum_{i=1}^n a(x_i) \le \mathbb{E}[a(X)](1+\epsilon).$$

Solution. *By counting the number of times each x occurs in the summation $\sum_{i=1}^n a(x_i)$, we have*

$$\begin{aligned}
\sum_{i=1}^n a(x_i) &= \sum_{x \in \mathcal{X}} n_x(\mathbf{x}) a(x) \\
&\le \sum_{x \in \mathcal{X}} nP_X(x)(1+\epsilon) a(x) \\
&= n(1+\epsilon)\mathbb{E}[a(X)],
\end{aligned}$$

where the inequality uses the non-negativity of $a(\cdot)$ along with the definition of strong typicality. The desired upper bound follows by dividing both sides by n. The lower bound follows using the same steps.

(c) Show that for any sequence $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{T}_n(\epsilon)$, it holds that

$$H(X)(1-\epsilon) \le \frac{1}{n} \sum_{i=1}^n \log_2 \frac{1}{P_X(x_i)} \le H(X)(1+\epsilon).$$

Notice that this means that strongly typical sequences are also typical according to the definition in the lecture (up to the replacement of ϵ by $\epsilon H(X)$, which essentially changes nothing since ϵ can be chosen arbitrarily).

Solution. *Just set $a(x) = \log_2 \frac{1}{P_X(x)}$ in part (b), which gives $\mathbb{E}[a(X)] = H(X)$.*

(d) Show that the the size of the typical set satisfies

$$2^{nH(X)(1-\epsilon)}(1-o(1)) \leq |\mathcal{T}_n(\epsilon)| \leq 2^{nH(X)(1+\epsilon)}$$

where $o(1)$ is a quantity that tends to zero as $n \to \infty$.

Solution. *Same as in the lecture once part (c) is established.*

3. **[Variable-Length Block Coding]**

Consider an n-bit string $\mathbf{x} \in \{0,1\}^n$ in which the all-zero string is chosen with probability $\mathbb{P}[\mathbf{x} = 00\ldots0] = \frac{1}{4}$, whereas with probability $\frac{3}{4}$, one of the remaining strings (not all zeros) is chosen uniformly at random.

(a) Give a simple scheme to perform lossless variable-length compression of such strings.
(Hint: Nothing fancy is needed (e.g., no need for typical sets).)

Solution. *The string* $00\ldots0$ *has a very high probability of occurring compared to the others, so it should be compressed as short as possible. We also need to add an initial flag bit to indicate whether the string is* x *or not.*

Therefore, we compress the all-zero string $00\ldots0$ *as* 0*, and for the others, prepend the flag bit* 1 *followed by the original string without compression (due to uniformity, it would be futile to try to compress further). The resulting scheme is lossless and uniquely decodeable.*

(b) What is the average length of the compressed string?

Solution. *The average length is*

$$\frac{1}{4}\cdot 1 + \frac{3}{4}\cdot(n+1) = \frac{3n}{4} + 1\,.$$

(c) How good is your scheme with respect to the Shannon entropy?

Solution. *The Shannon entropy of the ensemble of the compressed strings* X

$$H(X) \;=\; \frac{1}{4}\log_2 4 \;+\; (2^n-1)\,\frac{3}{4\,(2^n-1)}\,\log_2((4/3)\,(2^n-1)) \;=\; 0.8112 + \frac{3}{4}\log_2(2^n-1)\,.$$

For large n*, the dominating term both here and in part (b) is* $\approx \frac{3n}{4}$*, which shows that the compression scheme is quite effective.*

4. **[Asymptotic Equipartition Principle]**

Consider $X_1, X_2, \ldots, X_n, \ldots$, an infinite sequence iid random variables, each with probability distribution P_X. Let $\mathbf{X} = (X_1, \ldots, X_n)$, and let its (joint) distribution be $P_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$.

(a) Find an expression for $\lim_{n\to\infty} P_{\mathbf{X}}(\mathbf{x})^{\frac{1}{n}}$ that holds with high probability.

Solution. *This is essentially the Asymptotic Equipartition Principle (AEP).*

The weak law of large numbers states that

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X]\right| > \varepsilon\right] = 0$$

By the i.i.d. assumption,

$$\log_2 P_{\mathbf{X}}(\mathbf{x}) = \log_2 \prod_{i=1}^n P_X(X_i) = \sum_{i=1}^n \log_2 P_X(X_i)\,,$$

and since the random variable X's are i.i.d., so are the random variables $\log_2 P_X(X_i)$*. Therefore, by the law of large numbers,*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\log_2 P_X(X_i) - \mathbb{E}[\log_2 P_X(x)]\right| > \varepsilon\right] \;=\; 0$$

$$\Leftrightarrow \lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}\log_2 P_{\mathbf{X}}(\mathbf{X}) - (-H(X))\right| > \varepsilon\right] \;=\; 0$$

Equivalently, with high probability

$$\lim_{n\to\infty}\frac{1}{n}\log_2 P_{\mathbf{X}}(\mathbf{X}) = -H(X)\,.$$

and hence

$$\lim_{n\to\infty} P_{\mathbf{X}}(\mathbf{X})^{\frac{1}{n}} = 2^{-H(X)}\,.$$

since $a\log_2 b = \log_2 b^a$*.*

(b) Let $f(x)$ be an arbitrary function from $\mathcal{X}$ to the interval $(0,1]$. Find an expression for

$$\lim_{n\to\infty}\left[\prod_{i=1}^{n} f(X_i)\right]^{\frac{1}{n}}$$

that holds with high probability.

Solution. *Similar as in (a), we have*

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\log f(X_i) - \mathbb{E}[\log f(X)]\right| > \varepsilon\right) = 0\,.$$

Equivalently, with high probability

$$\lim_{n\to\infty}\frac{1}{n}\log\prod_{i=1}^{n} f(X_i) = \;\mathbb{E}[\log f(X)],$$

and hence

$$\lim_{n\to\infty}\left[\prod_{i=1}^{n} f(X_i)\right]^{\frac{1}{n}} = 2^{\mathbb{E}[\log f(X)]}\,.$$

5. **(Advanced) [Weighted Source Coding]**

In class, we saw that the minimum rate of compression for an i.i.d. source $\mathbf{X} = (X_1,\ldots,X_n)$ i.i.d. on a fixed distribution P_X is

$$H(X) = \sum_{x\in\mathcal{X}} P_X(x)\log_2\frac{1}{P_X(x)}.$$

Now suppose that there are **costs** to encoding each symbol. Consider a non-negative valued cost function $c(x)$. For any length-n string, let the string's cost be the product of individual costs:

$$c^{(n)}(\mathbf{x}) := \prod_{i=1}^{n} c(x_i).$$

Suppose that like with source coding, we assign unique index to some subset $\mathcal{A}_n \subseteq \mathcal{X}^n$. If any $\mathbf{X}$ is observed that we didn't assign an index to, then an error occurs, so the error probability is

$$\Pr(\text{err}) = \mathbb{P}[\mathbf{X} \notin \mathcal{A}_n]. \tag{2}$$

In addition, assigning $\mathbf{x}$ an index incurs a cost of $c^{(n)}(\mathbf{x})$, so some are more costly than others, and the total cost is

$$c^{(n)}(\mathcal{A}) := \sum_{\mathbf{x} \in \mathcal{A}_n} c^{(n)}(\mathbf{x}).$$

Setting all $c(x) = 1$ gives all $c^{(n)}(\mathbf{x}) = 1$ and recovers a total cost of $\mathcal{A}_n$ precisely equal to the size $|\mathcal{A}_n|$ (which we want to keep low in standard source coding).

We would like to know how low we can make $c^{(n)}(\mathcal{A}_n)$ while still ensuring that $\Pr(\text{err}) \to 0$ as $n \to \infty$. The answer turns out to be expressed in terms of the quantity

$$H(P\|c) := \sum_{x \in \mathcal{X}} P_X(x) \log_2 \frac{c(x)}{P_X(x)}.$$

(a) When $c(x) = 1$ for all $x \in \mathcal{X}$, we can bring $c^{(n)}(\mathcal{A}_n)$ down to $2^{n(R^*+\epsilon)}$ for small $\epsilon > 0$, but not down to $2^{n(R^*-\epsilon)}$. What is the value of R^*?

Solution. *The answer is $H(X)$, the entropy. (This is just the standard source coding theorem)*

(b) For a small $\epsilon > 0$, define the "typical set"

$$B_\epsilon^{(n)}(c) := \left\{ \mathbf{x} \,:\, H(P\|c) - \epsilon \le \frac{1}{n} \log_2 \frac{c^{(n)}(\mathbf{x})}{P_{\mathbf{X}}(\mathbf{x})} \le H(P\|c) + \epsilon \right\}.$$

For a general non-negative valued function $c(x)$, show that

$$\mathbb{P}[\mathbf{X} \in \mathcal{B}_\epsilon^{(n)}(c)] \to 1, \quad \text{as} \quad n \to \infty.$$

Solution. *Using the fact that both $c^{(n)}(\mathbf{x})$ and $P_{\mathbf{X}}(\mathbf{x})$ factorize as a product from $i = 1$ to n, we have*

$$\begin{aligned}
\mathbb{P}[\mathbf{X} \in \mathcal{B}_\epsilon^{(n)}(c)] &= \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n \log_2 \frac{c(X_i)}{P_X(X_i)} - H(P\|c)\right| \le \epsilon\right] \\
&= 1 - \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n \log_2 \frac{c(X_i)}{P_X(X_i)} - H(P\|c)\right| > \epsilon\right] \\
&\to 1
\end{aligned}$$

by the law of large numbers.

(c) Show that

$$c^{(n)}\left(\mathcal{B}_\epsilon^{(n)}(c)\right) \le 2^{n(H(P\|c)+\epsilon)}.$$

(Hint: Write the left-hand side as a sum of $c^{(n)}(\mathbf{x})$ values, and upper bound each $c^{(n)}(\mathbf{x})$ using the definition of $\mathcal{B}_\epsilon^{(n)}(c)$.)

Solution. *We have*

$$\begin{aligned}
c^{(n)}\left(\mathcal{B}_\epsilon^{(n)}(c)\right) &= \sum_{\mathbf{x}\in\mathcal{B}_\epsilon^{(n)}(c)} c^{(n)}(\mathbf{x}) \\
&\overset{(a)}{\le} \sum_{\mathbf{x}\in\mathcal{B}_\epsilon^{(n)}(c)} P_{\mathbf{X}}(\mathbf{x})2^{n(H(P\|c)+\epsilon)} \\
&= 2^{n(H(P\|c)+\epsilon)} \sum_{\mathbf{x}\in\mathcal{B}_\epsilon^{(n)}(c)} P_{\mathbf{X}}(\mathbf{x}) \\
&\overset{(b)}{\le} 2^{n(H(P\|c)+\epsilon)},
\end{aligned}$$

where (a) uses the simple re-arranging of the condition $\frac{1}{n}\log_2\frac{c^{(n)}(\mathbf{x})}{P_{\mathbf{X}}(\mathbf{x})} \le H(P\|c)+\epsilon$ *in the definition of* $B_\epsilon^{(n)}(c)$*, and (b) uses the fact that probabilities must always sum to at most one.*

(d) Using part (c), find some value R^* such that we can achieve $\Pr(\text{err}) \to 0$ with a total cost no higher than $2^{n(R^*+\epsilon)}$. (You do not need to show that this value of R^* is the best possible)

Solution. *By part (c), we can achieve* $R^* = H(P\|c)$*. Just set* $\mathcal{A}_n = \mathcal{B}_\epsilon^{(n)}(c)$*.*