CS3236: 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)

   (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.)

   (c) Your answer to part (b) should come down to showing that the proportion of heads (or tails) is close to the average. Your answer to part (a) could in fact also be interpreted similarly, but instead requiring the *sum* of the number of heads and tails to be close to its average (which is trivially true, since this sum is n).

   For sources with more than two symbols, it turns out that we shouldn't expect an equivalence between being in $\mathcal{T}_n(\epsilon)$ and certain counts (or sums of counts) being close to their average. Find an example that supports this claim.

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 $\epsilon > 0$.

   (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

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

   (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).

(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 $n \to \infty$.

3. **[Equality in Fano's Inequality]**

Fano's inequality states that for two random variables X and $\hat{X}$ on a common alphabet $\mathcal{X}$, it holds that

$$H(X|\hat{X}) \leq H_2(P_{\mathrm{e}}) + P_{\mathrm{e}} \log_2(|\mathcal{X}| - 1),$$

where $H_2(\cdot)$ is the binary entropy function, and $P_{\mathrm{e}} = \mathbb{P}[\hat{X} \neq X]$.

Suppose that $\mathcal{X} = \{1, 2, 3, 4, 5\}$ and that $P_{\mathrm{e}} = 0.2$. Describe (in words or mathematically) a joint distribution on $(X, \hat{X})$ that makes Fano's inequality hold with equality.

4. **[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).)*

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

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

5. **[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.

(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.

6. **(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

$$\Pr(\mathrm{err}) = \mathbb{P}[\mathbf{X} \notin \mathcal{A}_n]. \tag{1}$$

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^*?

(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

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

(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 $\mathcal{B}_\epsilon^{(n)}(c)$.)

(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)

Hints

1. (a) is straightforward because all sequences are equally likely. In (b) try to simplify the property in the typical set's definition by taking the log and simplifying.

2. In (a) use the law of large numbers and the union bound. In (b) write the summation in terms of $n_x(\mathbf{x})$ and then apply the bounds in the definition of $\mathcal{T}_n(\epsilon)$. (c) is a special case of part b.

3. A simple strategy is to just assign a very short (1-bit) string to probability-$\frac{1}{4}$ sequence, and then assign everything else a unique of a fixed length.

4. In (a), apply $2^{(\cdot)}$ to both sides of the equation $\lim_{n\to\infty} \frac{1}{n} \log_2 P_{\mathbf{X}}(\mathbf{X}) = -H(X)$. In (b) argue similarly with $f(\cdot)$ replacing $P_X(\cdot)$.

5. (a) just reduces to something we are already familiar with. In (b) expand out $c^{(n)}(\mathbf{x})$ and $P_{\mathbf{X}}(\mathbf{x})$ and use the law of large numbers. In (c) use the hint given. (d) is a one-line answer given part c.