Problem Set 2 — *Due Friday, October 25, before class starts*

For the Exercise Sessions on Oct 14 and 18

Last name	First name	SCIPER Nr	Points

Problem 1: Elias coding

Let 0^n denote a sequence of n zeros. Consider the code (the subscript U a mnemonic for 'Unary'), $\mathcal{C}_U : \{1, 2, \dots\} \to \{0, 1\}^*$ for the positive integers defined as $\mathcal{C}_U(n) = 0^{n-1}$.

(a) Is $\mathcal{C}_U$ injective? Is it prefix-free?

Consider the code (the subscript B a mnenonic for 'Binary'), $\mathcal{C}_B : \{1, 2, \dots\} \to \{0, 1\}^*$ where $\mathcal{C}_B(n)$ is the binary expansion of n. I.e., $\mathcal{C}_B(1) = 1$, $\mathcal{C}_B(2) = 10$, $\mathcal{C}_B(3) = 11$, $\mathcal{C}_B(4) = 100$, ... Note that

$$\text{length}\, \mathcal{C}_B(n) = \lceil \log_2(n + 1) \rceil = 1 + \lfloor \log_2 n \rfloor.$$

(b) Is $\mathcal{C}_B$ injective? Is it prefix-free?

With $k(n) = \text{length}\, \mathcal{C}_B(n)$, define $\mathcal{C}_0(n) = \mathcal{C}_U(k(n))\mathcal{C}_B(n)$.

(c) Show that $\mathcal{C}_0$ is a prefix-free code for the positive integers. To do so, you may find it easier to describe how you would recover $n_1, n_2, \dots$ from the concatenation of their codewords $\mathcal{C}_0(n_1)\mathcal{C}_0(n_2)\dots$.

(d) What is $\text{length}(\mathcal{C}_0(n))$?

Now consider $\mathcal{C}_1(n) = \mathcal{C}_0(k(n))\mathcal{C}_B(n)$.

(e) Show that $\mathcal{C}_1$ is a prefix-free code for the positive integers, and show that $\text{length}(\mathcal{C}_1(n)) = 2 + 2\lfloor \log(1 + \lfloor \log n \rfloor) \rfloor + \lfloor \log n \rfloor \le 2 + 2\log(1 + \log n) + \log n$.

Suppose U is a random variable taking values in the positive integers with $\Pr(U = 1) \ge \Pr(U = 2) \ge \dots$.

(f) Show that $E[\log U] \le H(U)$, [Hint: first show $i \Pr(U = i) \le 1$], and conclude that

$$E[\text{length}\, \mathcal{C}_1(U)] \le H(U) + 2\log(1 + H(U)) + 2.$$

Solution

(a) As $\mathcal{C}_U(n)$ and $\mathcal{C}_U(m)$ are of different lengths when $n \neq m$, the code is injective. It is not prefix free, in particular $\mathcal{C}_U(1) =$ empty-string is a prefix of all other codewords.

(b) As different integers have different binary expansions, $\mathcal{C}_B$ is injective. It is not prefix free, e.g., $\mathcal{C}_B(1) = 1$ is a prefix of all other codewords.

(c) The codeword of $\mathcal{C}_0(n) = \mathcal{C}_U(k(n))\mathcal{C}_B(n)$ is concatenated by two parts. The first part, $\mathcal{C}_U(k(n))$, is the sequence of zeros with length of $k(n)-1$. And the second part, $\mathcal{C}_B(n)$ is a binary representation for n. For any two different positive integers n_1 and n_2, let's assume that $n_1 < n_2$, which implies that $\text{length}(\mathcal{C}_0(n_1)) \leq \text{length}(\mathcal{C}_0(n_2))$ and $k(n_1) \leq k(n_2)$. We show that $\mathcal{C}_0(n_1)$ is not a prefix of $\mathcal{C}_0(n_2)$.

If $k(n_1) < k(n_2)$, the first $k(n_1)$ bits of $\mathcal{C}_0(n_1)$ are $0\dots01$[1], while the first $k(n_1)$ bits of $\mathcal{C}_0(n_2)$ are all zeros. So in such cases, $\mathcal{C}_0(n_1)$ cannot be a prefix of $\mathcal{C}_0(n_2)$. If $k(n_1) = k(n_2)$, we have $\text{length}(\mathcal{C}_0(n_1)) = \text{length}(\mathcal{C}_0(n_2))$. Although the first $k(n_1)$ bits of $\mathcal{C}_0(n_1)$ and $\mathcal{C}_0(n_2)$ are the same, the second parts, $\mathcal{C}_B(n_1)$ and $\mathcal{C}_B(n_2)$ are different. So $\mathcal{C}_0(n_1)$ cannot be a prefix of $\mathcal{C}_0(n_2)$. Therefore, $\mathcal{C}_0(n_1)$ cannot be a prefix of $\mathcal{C}_0(n_2)$ for any positive integers $n_1 < n_2$. In other words, $\mathcal{C}_0$ is a prefix-free code for the positive integers.

(d)Since $k(n) = \text{length}(\mathcal{C}_B(n)) = 1 + \lfloor \log_2 n \rfloor$,

$$\begin{aligned}
\text{length}(\mathcal{C}_0(n)) &= \text{length}(\mathcal{C}_U(k(n))) + \text{length}(\mathcal{C}_B(n)) \\
&= k(n) - 1 + 1 + \lfloor \log_2 n \rfloor \\
&= 1 + 2\lfloor \log_2 n \rfloor
\end{aligned}$$

(e) Similarly, as we did in (c), we can show that for any positive integers $n_1 < n_2$, $\mathcal{C}_1(n_1)$ cannot be a prefix of $\mathcal{C}_1(n_2)$. If $k(n_1) < k(n_2)$, $\mathcal{C}_0(k(n_1))$ is not a prefix of $\mathcal{C}_0(k(n_2))$, since $\mathcal{C}_0$ is prefix-free for positive integers. Hence, in such cases, $\mathcal{C}_1(n_1)$ cannot be a prefix of $\mathcal{C}_1(n_2)$. If $k(n_1) = k(n_2)$, we have $\text{length}(\mathcal{C}_1(n_1)) = \text{length}(\mathcal{C}_1(n_2))$. Although the first $\text{length}(\mathcal{C}_0(k(n_1)))$ bits of $\mathcal{C}_1(n_1)$ and $\mathcal{C}_1(n_2)$ are the same, the second parts, $\mathcal{C}_B(n_1)$ and $\mathcal{C}_B(n_2)$ are different. So $\mathcal{C}_1(n_1)$ cannot be a prefix of $\mathcal{C}_1(n_2)$. Therefore, $\mathcal{C}_1(n_1)$ cannot be a prefix of $\mathcal{C}_1(n_2)$ for any positive integers $n_1 < n_2$. In other words, $\mathcal{C}_1$ is a prefix-free code for the positive integers.

The length of $\mathcal{C}_1(n)$ can be computed as

$$\begin{aligned}
\text{length}(\mathcal{C}_1(n)) &= \text{length}(\mathcal{C}_0(k(n))) + \text{length}(\mathcal{C}_B(n)) \\
&= 1 + 2\lfloor \log_2 k(n) \rfloor + k(n) \\
&= 2 + 2\lfloor \log_2(1 + \lfloor \log_2 n \rfloor) \rfloor + \lfloor \log_2 n \rfloor \\
&\leq 2 + 2\log_2(1 + \log_2 n) + \log_2 n
\end{aligned}$$

(f) For random variable U with $\Pr(U=1) \geq \Pr(U=2) \geq \dots$, we have

$$1 = \sum_j \Pr(U=j) \geq \sum_{j=1}^{i} \Pr(U=j) \geq i \Pr(U=i)$$

Taking log at both sides, we get $-\log \Pr(U=i) \geq \log i, \forall i$.

$$E[\log U] = \sum_i \Pr(U=i) \log i \leq -\sum_i \Pr(U=i) \log \Pr(U=i) = H(U)$$

[1] If $k(n_1) = 1$, then there is no zeros and sequence starts with 1.

Using the results from (e) we have

$$\begin{aligned} E[\text{length}(\mathcal{C}_1(U))] &\leq E[2 + 2\log(1 + \log U) + \log U] \\ &= 2 + 2E[\log(1 + \log U)] + E[\log U] \\ &\leq 2 + 2\log(1 + H(U)) + H(U) \end{aligned}$$

where we used $E[\log(x)] \leq \log(E[x])$ for the second term because $\log(x)$ is a concave and monotonically increasing function.

Problem 2: Universal codes

Suppose we have an alphabet $\mathcal{U}$, and let Π denote the set of distributions on $\mathcal{U}$. Suppose we are given a family of S of distributions on $\mathcal{U}$, i.e., $S \subset \Pi$. For now, assume that S is finite.

Define the distribution $Q_S \in \Pi$

$$Q_S(u) = Z^{-1} \max_{P \in S} P(u)$$

where the normalizing constant $Z = Z(S) = \sum_u \max_{P \in S} P(u)$ ensures that Q_S is a distribution.

(a) Show that $D(P\|Q) \le \log Z \le \log |S|$ for every $P \in S$.

(b) For any S, show that there is a prefix-free code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ such that for any random variable U with distribution $P \in S$,

$$E[\text{length}\, \mathcal{C}(U)] \le H(U) + \log Z + 1.$$

(Note that $\mathcal{C}$ is designed on the knowledge of S alone, it cannot change on the basis of the choice of P.) [Hint: consider $L(u) = -\log_2 Q_S(u)$ as an 'almost' length function.]

(c) Now suppose that S is not necessarily finite, but there is a finite $S_0 \subset \Pi$ such that for each $u \in \mathcal{U}$, $\sup_{P \in S} P(u) \le \max_{P \in S_0} P(u)$. Show that $Z(S) \le |S_0|$.

Now suppose $\mathcal{U} = \{0,1\}^m$. For $\theta \in [0,1]$ and $(x_1, \dots, x_m) \in \mathcal{U}$, let

$$P_\theta(x_1, \dots, x_n) = \prod_i \theta^{x_i}(1-\theta)^{1-x_i}.$$

(This is a fancy way to say that the random variable $U = (X_1, \dots, X_n)$ has i.i.d. Bernoulli θ components). Let $S = \{P_\theta : \theta \in [0,1]\}$.

(d) Show that for $u = (x_1, \dots, x_m) \in \{0,1\}^m$

$$\max_\theta P_\theta(x_1, \dots, x_m) = P_{k/m}(x_1, \dots, x_m)$$

where $k = \sum_i x_i$.

(e) Show that there is a prefix-free code $\mathcal{C} : \{0,1\}^m \to \{0,1\}^*$ such that whenever $X_1, \dots, X_n$ are i.i.d. Bernoulli,

$$\frac{1}{m} E[\text{length}\, \mathcal{C}(X_1, \dots, X_m)] \le H(X_1) + \frac{1 + \log_2(1+m)}{m}.$$

Solution

(a) From the definition $Q_S(u) = Z^{-1} \max_{P \in S} P(u)$, we have $Q_S(u) \ge P(u)/Z$. Hence, $Z \ge P(u)/Q_S(u)$ and

$$D(P\|Q) = \sum_u P(u) \log \frac{P(u)}{Q(u)} \le \sum_u P(u) \log Z = \log Z$$

From $Z = Z(S) = \sum_u \max_{P \in S} P(u)$, we have $Z \le \sum_u \sum_{P \in S} P(u) = \sum_{P \in S} \sum_u P(u) = |S|$. So $\log Z \le \log |S|$.

(b) For any S, we can find a binary code with length function $L(u) = \lceil -\log_2 Q_S(u) \rceil$ for the codeword $\mathcal{C}(u)$. Since the length function of this binary code satisfies the Kraft Inequality,

$$\sum_u 2^{-L(u)} = \sum_u 2^{-\lceil -\log_2 Q_S(u) \rceil} \le \sum_u 2^{\log_2 Q_S(u)} \le \sum_u Q_S(u) = 1$$

there exists a prefix-free code $\mathcal{C}$ with length function $L(u)$. And the expected length of such code can be computed as

$$\begin{aligned} E[\text{length}\,\mathcal{C}(U)] = E[L(U)] &= E[\lceil -\log_2 Q_S(u) \rceil] \\ &\leq E[1 - \log_2 Q_S(u)] \\ &= 1 + E[\log_2 \frac{P(u)}{Q_S(u)} + \log_2 \frac{1}{P(u)}] \\ &= 1 + D(P \| Q) + H(U) \\ &\leq 1 + \log Z + H(U) \end{aligned}$$

(c) Similar as we showed in (a),

$$Z(S) = \sum_u \max_{P \in S} P(u) \leq \sum_u \sup_{P \in S} P(u) \leq \sum_u \max_{P \in S_0} P(u) \leq \sum_u \sum_{P \in S_0} P(u) = |S_0|$$

(d) Rewrite the definition of P_θ:

$$P_\theta(x_1, \ldots, x_m) = \prod_i \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_i x_i} (1-\theta)^{\sum_i (1-x_i)} = \theta^k (1-\theta)^{m-k}$$

Thus, $\log P_\theta = k \log \theta + (m-k) \log(1-\theta)$.

Compute the differentiation of $\log P_\theta$ w.r.t θ:

$$\frac{d}{d\theta} \log P_\theta = \frac{k}{\theta} - \frac{m-k}{1-\theta}$$

Set $\frac{d}{d\theta} \log P_\theta = 0$, we get $\hat{\theta} = k/m$. As logarithm is an increasing function, P_θ is maximized when $\log P_\theta$ is maximized.

(e) From (b) we know that there exists a prefix-free code such that

$$E[\text{length}\,\mathcal{C}(X_1, \ldots, X_m)] \leq H(X_1, \ldots, X_m) + \log Z + 1$$

where $H(X_1, \ldots, X_m) = mH(X_1)$, since they are i.i.d. From (d), we know that $S_0 = \{P_{k/m} : k = \sum_i^m x_i\}$ has the property in (c). Since each x_i is binary, k is an integer between 0 and m. So $|S_0| = m+1$, we have $Z(S) \leq |S_0| = m+1$. Therefore we have

$$\frac{1}{m} E[\text{length}\,\mathcal{C}(X_1, \ldots, X_m)] \leq H(X_1) + \frac{\log(1+m)+1}{m}$$

Problem 3: Prediction and coding

After observing a binary sequence $u_1, \ldots, u_i$, that contains $n_0(u^i)$ zeros and $n_1(u^i)$ ones, we are asked to estimate the probability that the next observation, u_{i+1} will be 0. One class of estimators are of the form

$$\hat{P}_{U_{i+1}|U^i}(0|u^i) = \frac{n_0(u^i) + \alpha}{n_0(u^i) + n_1(u^i) + 2\alpha} \quad \hat{P}_{U_{i+1}|U^i}(1|u^i) = \frac{n_1(u^i) + \alpha}{n_0(u^i) + n_1(u^i) + 2\alpha}.$$

We will consider the case $\alpha = 1/2$, this is known as the Krichevsky–Trofimov estimator. Note that for $i = 0$ we get $\hat{P}_{U_1}(0) = \hat{P}_{U_1}(1) = 1/2$.

Consider now the joint distribution $\hat{P}(u^n)$ on $\{0,1\}^n$ induced by this estimator,

$$\hat{P}(u^n) = \prod_{i=1}^{n} \hat{P}_{U_i|U^{i-1}}(u_i|u^{i-1}).$$

(a) Show, by induction on n that, for any n and any $u^n \in \{0,1\}^n$,

$$\hat{P}(u_1, \ldots, u_n) \geq \frac{1}{2\sqrt{n}} \left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1},$$

where $n_0 = n_0(u^n)$ and $n_1 = n_1(u^n)$.
[Hint: if $0 \leq m \leq n$, then $(1 + 1/n)^{n+1/2} \geq \frac{m+1}{m+1/2}(1 + 1/m)^m$]

(b) Conclude that there is a prefix-free code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ such that

$$\text{length}\, \mathcal{C}(u_1, \ldots, u_n) \leq n h_2\left(\frac{n_0(u^n)}{n}\right) + \frac{1}{2}\log n + 2,$$

with $h_2(x) = -x \log x - (1-x)\log(1-x)$.

(c) Show that if $U_1, \ldots, U_n$ are i.i.d. Bernoulli, then

$$\frac{1}{n} E[\text{length}\, \mathcal{C}(U_1, \ldots, U_n)] \leq H(U_1) + \frac{1}{2n}\log n + \frac{2}{n}$$

Solution

(a) For $n = 1$, we have $\hat{P}(u_1) = \hat{P}_{U_1}(u_i) = \frac{1}{2}$. If $u_1 = 0$, $n_0(u_1) = 1$ and $n_1(u_1) = 0$. Hence, $\hat{P}(u_1) = \frac{1}{2} = \frac{1}{2\sqrt{n}}(\frac{n_0}{n})^{n_0}(\frac{n_1}{n})^{n_1}$. It is easy to show that for $u_1 = 1$, the inequality still holds with equality.

For $n = k \geq 1$, let's assume that $\hat{P}(u_1, \ldots, u_k) \geq \frac{1}{2\sqrt{k}}\left(\frac{n_0}{k}\right)^{n_0}\left(\frac{n_1}{k}\right)^{n_1}$. For $n = k+1$, it is sufficient to check $u_{k+1} = 0$, as the case $u_{i+1} = 1$ is the same if we also exchange the roles of n_0 and n_1. In this case, $n_0(u^{k+1}) = n_0(u^k) + 1$ and $n_1(u^{k+1}) = n_1(u^k)$.

$$\begin{aligned}
\hat{P}(u_1, \ldots, u_k, 0) &= \hat{P}_{U_{k+1}|U^k}(0|u^k)\hat{P}_{U^k}(u^k) \\
&\geq \frac{n_0(u^k) + \frac{1}{2}}{n_0(u^k) + n_1(u^k) + 1} \frac{1}{2\sqrt{k}} \left(\frac{n_0(u^k)}{k}\right)^{n_0(u^k)} \left(\frac{n_1(u^k)}{k}\right)^{n_1(u^k)} \\
&= \underbrace{\frac{(k+1)^{k+1/2}}{k^{k+1/2}} \frac{(n_0(u^k) + \frac{1}{2}) n_0(u^k)^{n_0(u^k)}}{(n_0(u^k)+1)^{n_0(u^k)+1}}}_{f(u^k)} \frac{1}{2\sqrt{k+1}} \left(\frac{n_0(u^{k+1})}{k+1}\right)^{n_0(u^{k+1})} \left(\frac{n_1(u^{k+1})}{k+1}\right)^{n_1(u^{k+1})}
\end{aligned}$$

We need to show that $f(u^k) \geq 1$ for any $u^k \in \{0,1\}^k$, but this follows from the hint. Therefore, we proved that our induction hypothesis is true for any $n = k+1$, given the condition that $n = k$ cases is satisfied. By induction, we have for any integer $n \geq 1$

$$\hat{P}(u_1, \ldots, u_n) \geq \frac{1}{2\sqrt{n}} \left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1},$$

Proof the hint: We need to show that:

$$\left(1 + \frac{1}{k}\right)^{k+1/2} \geq \underbrace{\frac{n_0(u^k)+1}{n_0(u^k)+\frac{1}{2}} \left(1 + \frac{1}{n_0(u^k)}\right)^{n_0(u^k)}}_{g(n_0(u^k)) = g(n_0)}.$$

Now, consider the function $g(x) = \frac{x+1}{x+\frac{1}{2}}(1+\frac{1}{x})^x$ for $x \geq 1$. Since we have that $n_0(u^k) \leq k$, if $g(x)$ is an increasing function then we would have:

$$\begin{aligned} g(n_0(u^k)) \leq g(k) = \frac{k+1}{k+\frac{1}{2}}(1+\frac{1}{k})^k &= \frac{k+1}{(k+\frac{1}{2})\sqrt{1+\frac{1}{k}}}(1+\frac{1}{k})^{k+1/2} \\ &= \frac{\sqrt{k(k+1)}}{k+\frac{1}{2}}(1+\frac{1}{k})^{k+1/2} \\ &< \left(1+\frac{1}{k}\right)^{k+1/2}, \end{aligned}$$

and the result would follow (the last inequality is due to $\sqrt{k(k+1)} < \sqrt{k(k+1)+1/4} = k+1/2$). Hence, we just need to show that $g(x)$ is an increasing function, *i.e.* that $\frac{d}{dx}g(x) \geq 0$. A simple way of doing this is by showing that $\ln g(x)$ is an increasing function, which would then imply the result for $g(x)$. If we compute the differentiation of $\ln g(x)$, we get

$$\frac{d}{dx}\ln g(x) = \frac{1}{x+1} - \frac{1}{x+\frac{1}{2}} + \ln\left(1+\frac{1}{x}\right) - \frac{1}{x+1} = \ln(x+1) - \ln x - \frac{1}{x+\frac{1}{2}}$$

Now observe:

$$\ln(x+1) - \ln x = \int_x^{x+1} \frac{1}{u} du = \mathbb{E}\left[\frac{1}{U}\right],$$

where U is a unifom random variable between x and $x+1$. Also,

$$\frac{1}{x+1/2} = \frac{1}{\mathbb{E}[U]}.$$

Thus:

$$\frac{d}{dx}\ln g(x) = \mathbb{E}\left[\frac{1}{U}\right] - \frac{1}{\mathbb{E}[U]}$$

and the positivity of $\frac{d}{dx}\ln g(x)$ follows from the convexity of the function $u \to 1/u$ (and Jensen's inequality).

(b) Consider the code with length function $L(u^n) = \lceil -\log \hat{P}(u^n) \rceil$. We can check that such code satisfies the Kraft Inequity.

$$\sum_{u^n} 2^{-L(u^n)} = \sum_{u^n} 2^{-\lceil -\log \hat{P}(u^n) \rceil} \leq \sum_{u^n} \hat{P}(u^n) = 1$$

Hence, there exists a prefix-free code with length function $L(u^n)$.

$$\begin{aligned}
\text{length}\,\mathcal{C}(u_1,\ldots,u_n) = \lceil -\log \hat{P}(u^n) \rceil &\le -\log \hat{P}(u^n) + 1 \\
&\le -\log\left(\frac{1}{2\sqrt{n}} \left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1} \right) + 1 \\
&= 2 + \frac{1}{2}\log n + n\left[-\frac{n_0}{n}\log(\frac{n_0}{n}) - \frac{n_1}{n}\log\frac{n_1}{n} \right] \\
&= 2 + \frac{1}{2}\log n + nh_2(\frac{n_0}{n})
\end{aligned}$$

(c) Let $\Pr(U_i = 0) = \theta$, $\forall i \in \{1,\ldots,n\}$. Since $U_1,\ldots,U_n$ are i.i.d, we have $E[n_0(u^n)] = \sum_{i=1}^{n} E[n_0(u_i)] = n\theta$ and $H(U_i) = h_2(\theta)$ for all i.

$$\begin{aligned}
E[\text{length}\,\mathcal{C}(U_1,\ldots,U_n)] &\le E[nh_2(\frac{n_0(u^n)}{n}) + \frac{1}{2}\log n + 2] \\
&= nE[h_2(\frac{n_0(u^n)}{n})] + \frac{1}{2}\log n + 2 \\
&\le nh_2(\frac{E[n_0(u^n)]}{n}) + \frac{1}{2}\log n + 2 \\
&= nh_2(\theta) + \frac{1}{2}\log n + 2 \\
&= nH(U_1) + \frac{1}{2}\log n + 2
\end{aligned}$$

Therefore,

$$\frac{1}{n}E[\text{length}\,\mathcal{C}(U_1,\ldots,U_n)] \le H(U_1) + \frac{1}{2n}\log n + \frac{2}{n}$$

Problem 4: Lower bound on Expected Length

Suppose U is a random variable taking values in $\{1, 2, \dots\}$. Set $L = \lfloor \log_2 U \rfloor$. (I.e., $L = j$ if and only if $2^j \le U < 2^{j+1}$; $j = 0, 1, 2, \dots$.

(a) Show that $H(U|L = j) \le j$, $j = 0, 1, \dots$.

(b) Show that $H(U|L) \le E[L]$.

(c) Show that $H(U) \le E[L] + H(L)$.

(d) Suppose that $\Pr(U = 1) \ge \Pr(U = 2) \ge \dots$. Show that $1 \ge i \Pr(U = i)$.

(e) With U as in (d), and using the result of (d), show that $E[\log_2 U] \le H(U)$ and conclude that $E[L] \le H(U)$.

(f) Suppose that N is a random variable taking values in $\{0, 1, \dots\}$ with distribution p_N and $E[N] = \mu$. Let G be a geometric random variable with mean μ, i.e., $p_G(n) = \mu^n/(1+\mu)^{1+n}$, $n \ge 0$.

Show that $H(G) - H(N) = D(p_N \| p_G)$, and conclude that $H(N) \le g(\mu)$ with $g(x) = (1 + x)\log_2(1 + x) - x \log_2 x$.

[Hint: Let $f(n, \mu) = -\log_2 p_G(n) = (n + 1)\log_2(1 + \mu) - n \log_2(\mu)$. First show that $E[f(G, \mu)] = E[f(N, \mu)]$, and consequently $H(G) = \sum_n p_N(n) \log_2(1/p_G(n))$.]

(g) Show that for U as in (d) and $g(x)$ as in (f),

$$E[L] \ge H(U) - g(H(U)).$$

[Hint: combine (f), (e), (c).]

(h) Now suppose U is a random variable taking values on an alphabet $\mathcal{U}$, and $c : \mathcal{U} \to \{0, 1\}^*$ is an injective code. Show that

$$E[\text{length}\, c(U)] \ge H(U) - g(H(U)).$$

[Hint: the best injective code will label $\mathcal{U} = \{a_1, a_2, a_3, \dots\}$ so that $\Pr(U = a_1) \ge \Pr(U = a_2) \ge \dots$, and assign the binary sequences $\lambda, 0, 1, 00, 01, 10, 11, ...$ to the letters $a_1, a_2, \dots$ in that order. Now observe that the i'th binary sequence in the list $\lambda, 0, 1, 00, 01, \dots$ is of length $\lfloor \log_2 i \rfloor$.]

Solution

(a) We know that if $L = j$ then $2^j \le U < 2^{j+1}$, meaning that if $L = j$ then U can take at most $2^{j+1} - 2^j = 2^j$ values. We also know that the entropy of a discrete random variable is at most the logarithm of the number of possible values it assumes. Thus,

$$H(U|L = j) \le \log_2(2^j) = j. \tag{1}$$

(b) We have that:

$$\begin{aligned} H(U|L) &= \sum_j p_L(j) H(U|L = j) && (2) \\ &\le \sum_j p_L(j) j && (3) \\ &= \mathbb{E}[L]. && (4) \end{aligned}$$

(c) We have that:

$$\begin{aligned} H(U) &\leq H(UL) && (5)\\ &= H(L) + H(U|L) && (6)\\ &\leq H(L) + \mathbb{E}[L]. && (7) \end{aligned}$$

Where (7) follows from *(b)*. Notice that Ineq. (5) is actually an equality, since L is a function of U (and thus, $H(L|U) = 0$).

(d) For random variable U with $\Pr(U = 1) \geq \Pr(U = 2) \geq \dots$, we have

$$1 = \sum_j \Pr(U = j) \geq \sum_{j=1}^{i} \Pr(U = j) \geq i \Pr(U = i). \tag{8}$$

(e) From *(d)* we get that for a given i, $\log_2 i \leq -\log_2 \Pr(U = i)$. Thus:

$$\begin{aligned} \mathbb{E}[\lfloor \log_2 U \rfloor] &= \sum_i \Pr(U = i) \lfloor \log_2 i \rfloor && (9)\\ &\leq \sum_i \Pr(U = i) \log_2 i && (10)\\ &\leq -\sum_i \Pr(U = i) \log_2 \Pr(U = i) && (11)\\ &= H(U) && (12) \end{aligned}$$

(f) It is easy to see that, for any integer valued random variable Q:

$$\begin{aligned} \mathbb{E}[f(Q,\mu)] &= \sum_n ((n+1)\log(1+\mu) - n \log \mu) p_Q(n) && (13)\\ &= \log(1+\mu) \sum_n (n+1) p_Q(n) - \log \mu \sum_n n p_Q(n) && (14)\\ &= \log(1+\mu)(\mathbb{E}[Q] + 1) - \log \mu \mathbb{E}[Q] && (15) \end{aligned}$$

Thus, since $\mathbb{E}[N] = \mathbb{E}[G]$, we have that $\mathbb{E}[f(N,\mu)] = \mathbb{E}[f(G,\mu)]$.

This implies that $H(G) = \sum_n p_N(n) \log(1/p_G(n))$ as $H(G) = \mathbb{E}_G[-\log(p_G)] = \mathbb{E}_N[-\log(p_G)]$. Computing the difference:

$$\begin{aligned} H(G) - H(N) &= \sum_n p_N(n) \left(\log \frac{1}{p_G(n)} - \log \frac{1}{p_N(n)} \right) && (16)\\ &= \sum_n p_N(n) \log \left(\frac{p_N(n)}{p_G(n)} \right) && (17)\\ &= D(p_N \| p_G). && (18) \end{aligned}$$

To conclude:

$$H(N) = H(G) - D(p_N \| p_G) \leq H(G) = (1+\mu)\log(1+\mu) - \mu \log \mu = g(\mu). \tag{19}$$

(g) Let us denote with $\mu = \mathbb{E}[L]$. L takes values in $\{0, 1, \dots\}$ and from *(f)* we know that

$$H(L) \leq g(\mu). \tag{20}$$

From *(e)* we have that

$$\mu = \mathbb{E}[L] \leq H(U). \tag{21}$$

As $g(x)$ a non-decreasing function for $x > 0$ (the derivative is $\log_2(1+x) - \log_2(x) > 0$ for $x > 0$), we can see that

$$g(\mu) = g(\mathbb{E}[L]) \leq g(H(U)). \tag{22}$$

To conclude, from *(c)* we have that:

$$\begin{aligned} \mathbb{E}[L] &\geq H(U) - H(L) && (23)\\ &\geq H(U) - g(\mu) && (24)\\ &\geq H(U) - g(H(U)). && (25) \end{aligned}$$

(h) Consider the following random variable V taking values in the alphabet $\mathcal{V} = \{1, 2, \ldots\}$ and such that $\Pr(V = i) = \Pr(U = a_i)$ for every $i = 1, 2 \ldots,$ *i.e.* a bijective mapping from U to V. We have that $\mathbb{E}[\text{length } c(U)] = \mathbb{E}[\lfloor \log_2 V \rfloor]$. Let us denote with $\hat{L} = \lfloor \log_2 V \rfloor$: this random variable will play the same role played by L until now. We can say that:

$$\begin{aligned} \mathbb{E}[\text{length } c(U)] &= \mathbb{E}[\hat{L}] && (26)\\ &\geq H(V) - g(H(V)) && (27)\\ &= H(U) - g(H(U)). && (28) \end{aligned}$$

Where (27) follows from *(g)* and (28) is true since V is a bijective function of U and entropy is preserved under bijective mappings.

Problem 5: Code Extension

Suppose $|\mathcal{U}| \geq 2$. For $n \geq 1$ and a code $c : \mathcal{U} \to \{0,1\}^*$ we define its n*-extension* $c^n : \mathcal{U}^n \to \{0,1\}*$ via $c^n(u^n) = c(u_1)\dots c(u_n)$. In other words $c^n(u^n)$ is the concatenation of the binary strings $c(u_1)$, $\dots$, $c(u_n)$. A code c is said to be *uniquely decodeable* if for any u^k and $\tilde{u}^m$ with $u^k \neq \tilde{u}^m$, $c^k(u^k) \neq c^m(\tilde{u}^m)$.

(a) Show that if c is uniquely decodable, then for all $n \geq 1$, c^n is injective.

(b) Show that if c is not uniquely decodable, there are u^k and $\tilde{u}^m$ with $u_1 \neq \tilde{u}_1$ and $c^k(u^k) = c^m(\tilde{u}^m)$.

(c) Show that if c is not uniquely decodable, then there is an n for which c^n is not injective. [Hint: try $n = k + m$.]

Solution

(a) Suppose that c^n is not injective, then there exists $u^n \neq \tilde{u}^n$ such that $c^n(u^n) = c^n(\tilde{u}^n)$, hence c is not uniquely decodable, which is a contradiction.

(b) If c is not uniquely decodable, then there exists u^k and $\tilde{u}^m$ such that $c^k(u^k) = c^m(\tilde{u}^m)$. First suppose that u^k is a prefix of $\tilde{u}^m$, then $c(\tilde{u}_{k+1}) = \lambda$ which means that for any $a \in \mathcal{U} \setminus \{\tilde{u}_{k+1}\}$ we have that $c^2(\tilde{u}_{k+1}a) = c^2(a\tilde{u}_{k+1})$ which proves the statement. If $\tilde{u}^m$ is a prefix of u^k a similar reasoning can be applied. Otherwise let p be the first index where $u_p \neq \tilde{u}_p$, then if $u_1^{p-1} = u_1 u_2 \dots u_{p-1}$, $u_p^k = u_p u_{p+1} \dots u_k$ and $\tilde{u}_p^m = \tilde{u}_p \tilde{u}_{p+1} \dots \tilde{u}_m$ we have that

$$c^{p-1}(u_1^{p-1})c^{k-p+1}(u_p^k) = c^k(u^k) = c^m(\tilde{u}^m) = c^{p-1}(u_1^{p-1})c^{m-p+1}(\tilde{u}_p^m)$$

Hence $c^{k-p+1}(u_p^k) = c^{m-p+1}(\tilde{u}_p^m)$ and $u_p \neq \tilde{u}_p$ which proves the statement.

(c) As shown in subquestion b, if c is not uniquely decodable then there exists u^k and $\tilde{u}^m$ such that $u_1 \neq \tilde{u}_1$ and $c^k(u^k) = c^m(\tilde{u}^m)$, now if $n = m + k$, we have that $c^n(u^k\tilde{u}^m) = c^k(u^k)c^m(\tilde{u}^m) = c^m(\tilde{u}^m)c^k(u^k) = c^n(\tilde{u}^m u^k)$ and since $u_1 \neq \tilde{u}_1$, $u^k\tilde{u}^m \neq \tilde{u}^m u^k$ so c^n is not injective.