Ecole Polytechnique Fédérale, Lausanne: Fall 2018
November 2, 2018

Problem Set 4 — *Due Friday, November 16, before class starts*

For the Exercise Sessions on Nov 2 and 9

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

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

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

With $k(n) = \operatorname{length} \mathcal{C}_B(n)$, 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 $\operatorname{length}(\mathcal{C}_0(n))$?

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 $\operatorname{length}(\mathcal{C}_1(n)) = 2 + 2\lfloor \log(1 + \lfloor \log n \rfloor) \rfloor + \lfloor \log n \rfloor \le 2 + 2\log(1 + \log n) + \log n$.

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

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

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 $0\ldots01$ [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 $\Pr(U=1) \geq \Pr(U=2) \geq \ldots$, 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

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

where we used $E[\log(x)] \le \log(E[x])$ 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 $(x_1, \ldots, x_m) \in \mathcal{U}$, let

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

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

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

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

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 $X_1, \ldots, X_n$ are i.i.d. Bernoulli,

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

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

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

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

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

(d) Rewrite the definition of P_θ:

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

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

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

where $H(X_1,\dots,X_m) = mH(X_1)$, 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 $Z(S) \le |S_0| = m+1$. Therefore we have

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

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 $u^n \in \{0,1\}^n$,

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

where $n_0 = n_0(u^n)$ and $n_1 = n_1(u^n)$.

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

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

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

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

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 $n_1(u^{k+1}) = n_1(u^k)$.

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

We need to show that $f(u^k) \geq 1$ for any $u^k \in \{0,1\}^k$. And it is equivalent to show that

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

Consider a continuous function $g(x) = \frac{x+1}{x+\frac{1}{2}}(1+\frac{1}{x})^x$ for $x \geq 1$, then $\ln g(x) = \ln(x+1) - \ln(x+\frac{1}{2}) + x\ln(1+\frac{1}{x})$. Compute the differentiation of $\ln g(x)$, we get

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

By taking derivative again, we get

$$\frac{d^2}{dx^2}\ln g(x) = \frac{d}{dx}\left[\ln(1+\frac{1}{x}) - \frac{1}{x+\frac{1}{2}}\right] = -\frac{1}{x(x+1)} + \frac{1}{(x+\frac{1}{2})^2} < 0, \forall x > 0$$

Hence, we have

$$\frac{d}{dx}\ln g(x) \geq \left[\ln(1+\frac{1}{x}) - \frac{1}{x+\frac{1}{2}}\right]_{x=\infty} = 0$$

Therefore, $\ln g(x)$ is increasing function for $x \geq 1$, which implies that $g(x)$ is also increasing function for $x \geq 1$. So, $g(n_0(u^k))$ is upper bounded by $g(k)$, when all bits are 0.

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

where the last inequality is due to $\sqrt{k(k+1)} < \sqrt{k(k+1)+1/4} = k+1/2$. 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 $n \geq 1$

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

(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 $L(u^n)$.

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

(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.

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

Therefore,

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

Problem 4: Lempel Ziv 78

Suppose $\dots, U_{-1}, U_0, U_1, \dots$ is a stationary process, i.e., for any $k = 1, 2, \dots$, any $u_0, \dots, u_{k-1}$, and any $n = \dots, -1, 0, 1, \dots$

$$\Pr(U_n \dots U_{n+k-1} = u_0 \dots u_{k-1}) = \Pr(U_0 \dots U_{k-1} = u_0 \dots u_{k-1}).$$

Suppose also that U is a recurrent process, i.e., any letter u_0 with $\Pr(U_0 = u_0) > 0$, the event $A = \{\text{there exists } i \geq 0 \text{ and } j > 0 \text{ such that } U_i = U_{-j} = u_0\}$ has $\Pr(A) = 1$. (That is, a positive probability letter u_0 will occur infinitely often.)

Fix u_0 with $\Pr(U_0 = u_0) > 0$. For $i \geq 0$ and $j < 0$, let

$$A_{ij} = \{U_i = u_0\} \cap \{U_{-j} = u_0\} \cap \bigcap_{k=-j+1}^{i-1} \{U_k \neq u_0\}$$

denote the event that j is the last time before time 0 that u_0 was seen and i was the first time after time 0 that u_0 is seen.

(a) Show that $\sum_{i \geq 0, j > 0} \Pr(A_{ij}) = \Pr(A) = 1$.

(b) Show that $\Pr(A_{ij}) = f(i + j)$, where

$$f(k) = \Pr(U_{-k} = u_0,\, U_{-l} \neq u_0 \text{ for } l = 1, \dots, k-1, U_0 = u_0).$$

(c) Using (a) and (b), show that

$$1 = \sum_{k \geq 1} k f(k) = 1.$$

(d) Let $K = \inf\{k > 0 : U_{-k} = u_0\}$ (i.e., the negative index of the most recent time before time 0 u_0 was seen). Observe that the event $\{K = k, U_0 = u_0\}$ is the event whose probability is $f(k)$. Using (c) show that

$$E[K | U_0 = u_0] = 1/\Pr(U_0 = u_0)$$

and that $E[\log K] \leq H(U_0)$.

Suppose we have a stationary and ergodic source $\dots, X_{-1}, X_0, X_1, \dots$. This means, in particular, that for any $n > 0$, the process $\{U_i\}$ defined by $U_i = (X_i, X_{i+1}, X_{i+n-1})$ is stationary and recurrent.

Fix a sequence $x_0, \dots, x_{n-1}$ with $\Pr\big((X_0 \dots X_{n-1}) = (x_0 \dots x_{n-1})\big) > 0$. Let

$$K = \inf\{k > 0 : (X_{-k} \dots X_{-k+n-1}) = (x_0 \dots x_{n-1})\}.$$

(e) Show that $E[\log K] \leq H(X_0 \dots X_{n-1})$.

(f) Consider the following data compression method. Assuming that the encoder has already described the infinite past $\dots, X_{-2}, X_{-1}$ to the decoder, he describes $X_0, \dots, X_{n-1}$ by (i) finding the most recent occurrence $X_0 \dots X_{n-1}$ in the past, (ii) describing the index K of this occurrence by the method of problem 1(f). Now that the decoder knows $\dots, X_{n-1}$, the encoder describes $X_n \dots X_{2n-1}$ is the same way, etc. Show that this method uses fewer than

$$\frac{1}{n} H(X_0 \dots X_{n-1}) + \frac{2}{n} \log(1 + H(X_0 \dots X_{n-1})) + \frac{2}{n}$$

bits per letter on the average.

Solution

(a) Note that $A = \cup A_{ij}$. As A_{ij} are disjoint, $\sum_{i\geq 0, j>0} \Pr(A_{ij}) = \Pr(\cup_{ij} A_{ij}) = \Pr(A) = 1$.

(b) By definition we have

$$\begin{aligned}
\Pr(A_{ij}) &= \Pr(U_{-j} = u_0, U_l \neq u_0 \text{ for } l = -j+1, \ldots, i-1, U_i = u_0) \\
&\overset{*}{=} \Pr(U_{-j-i} = u_0, U_{l-i} \neq u_0 \text{ for } l = -j+1, \ldots, i-1, U_{i-i} = u_0) \\
&= \Pr(U_{-j-i} = u_0, U_{-l} \neq u_0 \text{ for } l = 1, \ldots, i+j-1, U_0 = u_0) \\
&= f(i+j)
\end{aligned}$$

where $(*)$ is from stationary properties, $\Pr(U_{-j}^{i}) = \Pr(U_{-j-i}^{0})$.

(c) From (a) and (b), we have

$$1 = \sum_{i\geq 0, j>0} \Pr(A_{ij}) = \sum_{i\geq 0, j>0} f(i+j) \overset{k=i+j}{=} \sum_{k\geq 1} \sum_{0<j\leq k} f(k) = \sum_{k\geq 1} k f(k)$$

(d) Since $f(k) = \Pr(K = k, U_0 = u_0) = \Pr(U_0 = u_0) \Pr(K = k | U_0 = u_0)$, we have

$$E[K|U_0 = u_0] = \sum_k k \Pr(K = k|U_0 = u_0) = \sum_k k \frac{f(k)}{\Pr(U_0 = u_0)} = \frac{\sum_k k f(k)}{\Pr(U_0 = u_0)} = 1/\Pr(U_0 = u_0)$$

By Jensen's inequality,

$$E[\log K | U_0 = u_0] \leq \log E[K|U_0 = u_0] = -\log P(U_0 = u_0)$$

Taking expectation w.r.t U_0, we get

$$\begin{aligned}
E[\log K] &= E[E[\log K|U_0]] \\
&= E[-\log \Pr(U_0)] \\
&= H(U_0)
\end{aligned}$$

(e) Similar as we did in (d), with $U_0 = (X_0, \ldots, X_{n-1})$

$$\begin{aligned}
E[\log K] &= E[E[\log K|X_0, \ldots, X_{n-1}]] \\
&\leq E[-\log \Pr(X_0, \ldots, X_{n-1})] \\
&= H(X_0, \ldots, X_{n-1})
\end{aligned}$$

(f) We know from problem 1 that there is a code for positive integers in which K is represented by fewer than $2 + 2\log(1 + \log K) + \log K$ bits.

Since $E[\log K] \leq H(X_0, \ldots, X_{n-1})$, we know that

$$\begin{aligned}
E[\text{length}\, \mathcal{C}_1(K)] &\leq E[2 + 2\log(1 + \log K) + \log K] \\
&= 2 + 2E[\log(1 + \log K)] + E[\log K] \\
&\leq 2 + 2\log(1 + E[\log K]) + E[\log(K)] \\
&\leq 2 + 2\log(1 + H(X_0, \ldots, X_{n-1})) + H(X_0, \ldots, X_{n-1})
\end{aligned}$$

Since we have n letters, the average bits per letter is no larger than

$$\frac{2}{n} + \frac{2}{n}\log(1 + H(X_0, \ldots, X_{n-1})) + \frac{1}{n} H(X_0, \ldots, X_{n-1})$$

Problem 5: Quantization with two criteria

Suppose U^n has i.i.d. components with distribution P. We want to describe U^n at rate R, i.e., we want to design a function $f : \mathcal{U}^n \to \{1, \dots, 2^{nR}\}$.

We are given two distortion measures $d_1 : \mathcal{U} \times \mathcal{V}_1 \to \mathbb{R}$ and $d_2 : \mathcal{U} \times \mathcal{V}_2 \to \mathbb{R}$, and we wish to ensure that from $i = f(U^n)$ we can reconstruct $V_1^n = g_1(i) \in \mathcal{V}_1^n$ and $V_2^n = g_2(i) \in \mathcal{V}_2^n$ so that

$$E[d_1(U^n, V_1^n)] \le D_1 \quad \text{and} \quad E[d_2(U^n, V_2^n)] \le D_2$$

with given distortion criteria D_1 and D_2. (As in class $d(U^n, V^n) = \frac{1}{n}\sum_{i=1}^n d(U_i, V_i)$.)

(a) What is the rate distortion function $R(D_1, D_2)$?

(b) Suppose $R_1(D_1)$ is the rate distortion function with the first distortion criterion alone, and $R_2(D_2)$ is the rate distortion function with the second criterion alone. What relationship exists between $R(D_1, D_2)$ and $R_1(D_1) + R_2(D_2)$?

Solution

(a) We can define the rate distortion function as

$$R(D_1, D_2) = \min I(U; V_1 V_2) \text{ s.t. } E[d(U, V_1)] \le D_1, E[d(U, V_2)] \le D_2$$

(b) Suppose $p_1(v_1|u)$ achieves the minimum in $\min\{I(U;V_1) : E[d_1(U, V_1)] <= D_1\}$ and similarly for $p_2(v_2|u)$. Then, for the distribution $p(u, v_1, v_2) = p(u)p_1(v_1|u)p_2(v_2|u)$, the random variables V_1 and V_2 are conditionally independent given U, and thus $H(V_1V_2|U) = H(V_1|U) + H(V_2|U)$. Furthermore, under this distribution $E[d_1(U, V_1)] <= D_1$ and $E[d_2(U, V_2)] <= D_2$,

$$\begin{aligned} R(D_1, D_2) \le I(U; V_1V_2) &= H(V_1V_2) - H(V_1V_2|U) \\ &\le H(V_1) + H(V_2) - H(V_1|U) - H(V_2|U) \\ &= I(U; V_1) + I(U; V_2) \\ &= R_1(D_1) + R_2(D_2) \end{aligned}$$

We can also show that $R(D_1, D_2) \ge R_1(D_1)$ by noting that $I(V_1V_2; U) \ge I(V_1; U)$. Consequently, $R(D_1, D_2) \ge \max\{R_1(D_1), R_2(D_2)\}$.