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., $\mathcal{C}_B(1) = 1$, $\mathcal{C}_B(2) = 10$, $\mathcal{C}_B(3) = 11$, $\mathcal{C}_B(4) = 100$, $\dots$. 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

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

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 $D(P\|Q) \leq \log Z \leq \log |S|$ 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 $P \in S$,

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

(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 $u \in \mathcal{U}$, $\sup_{P \in S} P(u) \leq \max_{P \in S_0} P(u)$. Show that $Z(S) \leq |S_0|$.

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,

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

Problem 3: Prediction and coding

After observing a binary sequence $u_1, \dots, u_i$, 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 $u^n \in \{0,1\}^n$,

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

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

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

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

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

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

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

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 $g(x) = (1+x)\log(1+x) - x \log x$.

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

(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 $\lambda, 0, 1, 00, 01, 10, 11, \dots$ 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 $\lfloor \log_2 i \rfloor$.]

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 $n = k + m$.]