Problem Set 7 **(Graded)** —*Due Tuesday, Dec 19, before class starts*

For the Exercise Sessions on Dec 5 and 12

Last name	First name	SCIPER Nr	Points

Problem 1: Exponential Families and Maximum Entropy 1

Let $Y = X_1 + X_2$. Find the maximum entropy of Y under the constraint $\mathbb{E}[X_1^2] = P_1$, $\mathbb{E}[X_2^2] = P_2$:

(a) If X_1 and X_2 are independent.

(b) If X_1 and X_2 are allowed to be dependent.

Problem 2: Exponential Families and Maximum Entropy 2

Find the maximum entropy density f, defined for $x \geq 0$, satisfying $\mathbb{E}[X] = \alpha_1$, $\mathbb{E}[\ln X] = \alpha_2$. That is, maximize $-\int f \ln f$ subject to $\int x f(x) dx = \alpha_1$, $\int (\ln x) f(x) dx = \alpha_2$, where the integral is over $0 \leq x < \infty$. What family of densities is this?

Problem 3: Exponential Families and Maximum Entropy 3

For $t > 0$, consider a family of distributions supported on $[t, +\infty]$ such that $\mathbb{E}[\ln X] = \frac{1}{\alpha} + \ln t$, $\alpha > 0$.

1. What is the parametric form of a maximum entropy distribution satisfying the constraint on the support and the mean?
2. Find the exact form of the distribution.

Problem 4: Exponential Families and Maximum Entropy 4: I-projections

Let P denote the zero-mean and unit-variance Gaussian distribution. Assume that you are given N iid samples distributed according to P and let $\hat{P}_N$ be the empirical distribution.

Let Π denote the set of distributions with second moment $\mathbb{E}[X^2] = 2$. We are interested in

$$\lim_{N\to\infty} \frac{1}{N} \log \Pr\{\hat{P_N} \in \Pi\} = -\inf_{Q\in\Pi} D(Q\|P).$$

(a) Determine $-\text{arginf}_{Q\in\Pi} D(Q\|P)$, i.e., determine the element Q for which the infinum is taken on.

(b) Determine $-\inf_{Q\in\Pi} D(Q\|P)$.

The subsequent problems are to be solved during the exercise session on 12.12.23.

Problem 5: Choose the Shortest Description

Suppose $\mathcal{C}_0 : \mathcal{U} \to \{0,1\}^*$ and $\mathcal{C}_1 : \mathcal{U} \to \{0,1\}^*$ are two prefix-free codes for the alphabet $\mathcal{U}$. Consider the code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ defined by

$$\mathcal{C}(u) = \begin{cases} [0, \mathcal{C}_0(u)] & \text{if length}\mathcal{C}_0(u) \leq \text{length}\mathcal{C}_1(u) \\ [1, \mathcal{C}_1(u)] & \text{else.} \end{cases}$$

Observe that $\text{length}(\mathcal{C}(u)) = 1 + \min\{\text{length}(\mathcal{C}_0(u)), \text{length}(\mathcal{C}_1(u))\}$.

(a) Is $\mathcal{C}$ a prefix-free code? Explain.

(b) Suppose $\mathcal{C}_0, \dots, \mathcal{C}_{K-1}$ are K prefix-free codes for the alphabet $\mathcal{U}$. Show that there is a prefix-free code $\mathcal{C}$ with

$$\text{length}(\mathcal{C}(u)) = \lceil \log_2 K \rceil + \min_{0 \leq k < K-1} \text{length}(\mathcal{C}_k(u)).$$

(c) Suppose we are told that U is a random variable taking values in $\mathcal{U}$, and we are also told that the distribution p of U is one of K distributions $p_0, \dots, p_{K-1}$, but we do not know which. Using (b) describe how to construct a prefix-free code $\mathcal{C}$ such that

$$\mathbb{E}[\text{length}(\mathcal{C}(U))] \leq \lceil \log_2 K \rceil + 1 + H(U).$$

[Hint: From class we know that for each k there is a prefix-free code $\mathcal{C}_k$ that descibes each letter u with at most $\lceil -\log_2 p_k(u) \rceil$ bits.]

Problem 6: 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 $(x_1, \dots, x_m) \in \mathcal{U}$, 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,

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

Problem 7: 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 $\mathbb{E}[\log U] \le H(U)$, [Hint: first show $i \Pr(U = i) \le 1$], and conclude that

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