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.

Solution 1. *(a)* If X_1 and X_2 are independent,

$$\mathrm{Var}[Y] = \mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] \leq \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] = P_1 + P_2 \tag{1}$$

where equality holds when $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Thus we have

$$\max_{f(y)} h(Y) \quad \leq \quad \frac{1}{2}\log(2\pi e(P_1 + P_2)) \tag{2}$$

where equality holds when Y is Gaussian with zero mean, which requires X_1 and X_2 to be independent and Gaussian with zeros mean.

(b) For dependent X_1 and X_2, we have

$$\mathrm{Var}(Y) \leq \mathbb{E}[Y^2] = \mathbb{E}[(X_1 + X_2)^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\mathbb{E}[X_1 X_2] \leq (\sqrt{P_1} + \sqrt{P_2})^2 \tag{3}$$

where the first equality holds when $\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 0$, and the send equality holds when $X_2 = \sqrt{\frac{P_2}{P_1}} X_1$. Hence, $\max_{f(y)} h(Y) \leq \frac{1}{2}\log(2\pi e(\sqrt{P_1} + \sqrt{P_2})^2)$, where equality holds when Y is Gaussian with zero mean, which requires X_1 and X_2 to be Gaussian with zero mean and $X_2 = \sqrt{\frac{P_2}{P_1}} X_1$.

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?

Solution 2. The maximum entropy distribution subject to constraints

$$\int x f(x) dx = \alpha_1 \tag{4}$$

and

$$\int (\ln x) f(x) dx = \alpha_2 \tag{5}$$

is of the form

$$f(x) = e^{\lambda_0 + \lambda_1 x + \lambda_2 \ln x} = c x^{\lambda_2} e^{\lambda_1 x} \tag{6}$$

which is of the form of a Gamma distribution. The constants should be chosen so as to satisfy the constraints. We need to solve the following equations

$$\int_0^\infty f(x)dx = \int_0^\infty c x^{\lambda_2} e^{\lambda_1 x} dx = 1 \tag{7}$$

$$\int_0^\infty x f(x)dx = \int_0^\infty c x^{\lambda_2+1} e^{\lambda_1 x} dx = \alpha_1 \tag{8}$$

$$\int_0^\infty (\ln x) f(x)dx = \int_0^\infty c x^{\lambda_2} e^{\lambda_1 x} \ln x dx = \alpha_2 \tag{9}$$

Thus, the Gamma distributions $f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$ with

$$\mathbb{E}[X] = k\theta = \alpha_1 \qquad \mathbb{E}[\ln X] = \psi(k) + \ln(\theta) = \alpha_2 \tag{10}$$

is the exponential family we want.

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.

Solution 3. (i) The maximum entropy distribution has the parametric form $e^{\theta \ln x - A(\theta)} = x^\theta e^{-A(\theta)}$.

(ii) Let us first find the value of $A(\theta)$ from the density constraint $\int_t^\infty x^\theta e^{-A(\theta)} dx = 1$. This gives $e^{-A(\theta)} = -\frac{\theta+1}{t^{\theta+1}}$.

Next we find θ from the mean constraint $\int_t^\infty x^\theta e^{-A(\theta)} \ln x \, dx = \frac{1}{\alpha} + \ln t$. This gives $\frac{t^{\theta+1}((\theta+1)\ln t - 1)}{t^{\theta+1}(\theta+1)} = \ln t - \frac{1}{\theta+1} = \frac{1}{\alpha} + \ln t$ and therefore $\theta = -(\alpha+1)$. The resulting form of the distribution is

$$p(x) = \frac{\alpha t^\alpha}{x^{\alpha+1}}$$

.

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

Solution 4. We are looking for the I-projection of P onto Π, call the result Q. Since Π is a linear family with a single constraint on the expected value of x^2 we know that the density of the minimizing distribution has the form

$$q(x) = p(x)e^{\theta x^2 - A(\theta)}.$$

If we insert $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ this gives us

$$q(x) = e^{-\frac{x^2}{2} + \theta x^2 - \tilde{A}(\theta)}.$$

We recognize the right-hand side to be the density of a zero-mean Gaussian distribution and by assumption this distribution has second moment 2. Hence, the solution is a zero-mean Gaussian distribution with variance 2, i.e., $q(x) = \frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}}$. The asymptotic exponent is given by the KL distance between these two distributions. We have

$$\begin{aligned} D(q\|p) &= \int \frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}} \log \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} dx \\ &= \frac{1}{2} \log \frac{1}{2} + \int \frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}} [-\frac{x^2}{4} + \frac{x^2}{2}] dx \\ &= \frac{1}{2}(\log \frac{1}{2} + 1) = \frac{1}{2}(-\log 2 + 1) \sim 0.153426. \end{aligned}$$

To summarize

1. $-\text{arginf}_{Q \in \Pi} D(Q\|P)$ is given by $q(x) = \frac{1}{\sqrt{4\pi}} e^{-\frac{x^2}{4}}$.
2. $-\inf_{Q \in \Pi} D(Q\|P) = -0.153426$.

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 $\mathcal{C}_0, \ldots, \mathcal{C}_{K-1}$ 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 $p_0, \ldots, p_{K-1}$, 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.]

Solution 5. (a) Yes, $\mathcal{C}$ is a prefix-free code. We can prove it by contradiction. Suppose there exist $u, v \in \mathcal{U}$ such that $\mathcal{C}(u)$ is a prefix of $\mathcal{C}(v)$. Then they must start with the same bit. Without loss of generality, let us assume they start with 0, then we have $\mathcal{C}(u) = 0\mathcal{C}_0(u)$ is a prefix of $\mathcal{C}(v) = 0\mathcal{C}_0(v)$. This requires $\mathcal{C}_0(u)$ is a prefix of $\mathcal{C}_0(v)$ which contradicts to $\mathcal{C}_0$ is prefix free code.

(b) Generalizing the given construction, we can construct the code $\mathcal{C}(u)$ for any $u \in \mathcal{U}$ as follows.

$$\mathcal{C}(u) = \mathrm{Bin}(i^*)\mathcal{C}_{i^*}(u) \tag{11}$$

where $i^* = \arg\min_{0\le k \le K-1\}} \mathrm{length}\mathcal{C}_i(u)$ and $\mathrm{Bin}(i^*)$ is the binary representation of number i^*. The length of such code is exactly the given expression and by the same reason in (a), we can show that it is prefix-free.

(c) As the hint suggests, we can use prefix free code $\mathcal{C}_k$ such that $\mathrm{length}(\mathcal{C}_k) \le \lceil -\log_2 p_k(u) \rceil$ and construct the prefix-free code $\mathcal{C}$ as in [b]. Then we have

$$\begin{aligned}
\mathrm{length}(\mathcal{C}(u)) &= \lceil \log_2 K \rceil + \min_{0\le k<K-1} \mathrm{length}(\mathcal{C}_k(u)) && (12)\\
&\le \lceil \log_2 K \rceil + 1 - \min_{0\le k<K-1} \log_2 p_k(u) && (13)\\
&\le \lceil \log_2 K \rceil + 1 - \log_2 p(u) && (14)
\end{aligned}$$

Taking expectation at both sides, we get that

$$\mathbb{E}[\mathrm{length}(\mathcal{C}(U))] \le \lceil \log_2 K \rceil + 1 + H(U). \tag{15}$$

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

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

(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,

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

Solution 6. (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

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

(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:

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

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

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

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

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

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

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

Solution 7. (a) As $\mathcal{C}_U(n)$ and $\mathcal{C}_U(m)$ are of different lengths when $n \ne m$, 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 $-\log \Pr(U = i) \geq \log i, \forall i$.

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

Using the results from (e) we have

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

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

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