Homework 1 Solution and Grading Policy

TA: Kai-Chun Chen

Homework Policy: (Read before You Start to Work)

- Copying from other students' solution is not allowed. If caught, all involved students get 0 point on that particular homework. Caught twice, you will be asked to drop the course.
- Collaboration is welcome. You can work together with **at most one partner** on the homework problems which you find difficult. However, you should write down your own solution, not just copying from your partner's.
- Your partner should be the same for the entire homework.
- Put your collaborator's name beside the problems that you collaborate on.
- When citing known results from the assigned references, be as clear as possible.

1. (Phase transition in error probability) [20]

For a lossless source coding problem, let $\epsilon^*(n,k)$ denote the smallest possible error probability that any (n,k) source code can ever achieve, that is,

$$\epsilon^*(n,k) := \min\left\{\epsilon \,\middle|\, \text{there exists an } (n,k,\epsilon) \text{ source code}\right\}.$$

For a discrete memoryless source $\{S_i \mid i \in \mathbb{N}\}$, S_i's being i.i.d. copies of a discrete random variable S with entropy $\mathrm{H}(S)$, prove the following statements.

a) If $\mathrm{R} > \mathrm{H}(S)$, then
$$\lim_{n\to\infty} \epsilon^*(n, \lfloor n\mathrm{R}\rfloor) = 0 \qquad [10]$$

b) If $\mathrm{R} < \mathrm{H}(S)$, then
$$\lim_{n\to\infty} \epsilon^*(n, \lfloor n\mathrm{R}\rfloor) = 1. \qquad [10]$$

Note that these are alternative ways to state the achievability part and the converse part of the lossless source coding theorem in the lecture, respectively.

Solution:

1. a)

Claim: $\forall \epsilon \in (0,1), \exists N_0 \in \mathbb{N}$ such that a $(n, \lfloor nR \rfloor, \epsilon)$ code exists for all $n \geq N_0$.
Consider a δ-typical set $\mathcal{A}_\delta^{(n)}$ whose parameter $\delta > 0$ will be determined later. If we select this set to be the range of the decoding function, then the error probability is $\mathsf{Pr}\left\{S_n \notin \mathcal{A}_\delta^{(n)}\right\}$.
To make it a $(n, \lfloor nR \rfloor)$ code, we select $\delta \leq \frac{R-\mathrm{H}(S)}{2}$ so that

$$\lceil \log \left|\mathcal{A}_\delta^{(n)}\right| \rceil \leq n(\mathrm{H}(S) + \delta) + 1 \leq nR,$$

when n is large enough. By Proposition 1.2, we have:
For any $\epsilon \in (0,1), \exists N_0 \in \mathbb{N}$ such that $\mathsf{Pr}\left\{S_n \notin \mathcal{A}_\delta^{(n)}\right\} < \epsilon$ for all $n \geq N_0$. This perfectly proves the claim.
Therefore, for any $\epsilon \in (0,1)$, exist N_0 such that an (n, k, ϵ) code exists for all $n \geq N_0$. We can then conclude that $\lim\limits_{n\to\infty} \epsilon^*(n,k) = 0$.

Grading Policy

(10) Flawless proof
(8) Reasonable procedure but with some missing detail, e.g. undetermined δ
(6) Conceptually correct but lack of detail
(4) Conceptually correct but no explaination
(2) Wrong but I can tell you have tried

1. b)

Claim: $\forall \epsilon \in (0,1), \exists N_0 \in \mathbb{N}$ such that $\forall (n,k)$ code with $n \geq N_0$, $\mathsf{P}_\mathsf{e}^{(n)} > 1 - \epsilon$.
For any (n,k) code, denote its range of decoding function by $\mathcal{B}_n$.
Then we have for any $\delta' \in (0,1)$,

$$\begin{aligned}
\mathsf{Pr}\left\{\hat{S}_n = S_n\right\} &\leq \mathsf{Pr}\left\{S_n \in \mathcal{B}_n\right\} \\
&= \mathsf{Pr}\left\{S_n \in \mathcal{B}_n \cap \mathcal{A}_{\delta'}^{(n)}\right\} + \mathsf{Pr}\left\{S_n \in \mathcal{B}_n \cap \mathcal{A}_{\delta'}^{(n)^c}\right\} && \text{(by total probability)} \\
&\leq |B_n| \cdot 2^{-n(\mathrm{H}(S)-\delta')} + \mathsf{Pr}\left\{S_n \in \mathcal{A}_{\delta'}^{(n)^c}\right\} && \text{(by Proposition 1.1)} \\
&\leq 2^{nR} \cdot 2^{-n(\mathrm{H}(S)-\delta')} + \mathsf{Pr}\left\{S_n \in \mathcal{A}_{\delta'}^{(n)^c}\right\} \\
&= 2^{n(\mathrm{H}(S)-\delta)} \cdot 2^{-n(\mathrm{H}(S)-\delta')} + \mathsf{Pr}\left\{S_n \in \mathcal{A}_{\delta'}^{(n)^c}\right\} && \text{(say } R = \mathrm{H}(S) - \delta\text{)} \\
&= 2^{n(\delta'-\delta)} + \mathsf{Pr}\left\{S_n \in \mathcal{A}_{\delta'}^{(n)^c}\right\}
\end{aligned}$$

Now, we aim to select N_0 according to any given $\epsilon \in (0,1)$.
First, we choose $\delta' = \frac{\delta}{2}$, and we will have $N_1 = \frac{2}{\delta}\log\frac{2}{\epsilon}$ so that $2^{(\delta'-\delta)} < \frac{\epsilon}{2}$, $\forall n \ge N_1$.
Next, by Proposition 1.2, there exists $N_2 \in \mathbb{N}$ such that $\mathsf{Pr}\left\{S_n \in \mathcal{A}_{\frac{\delta}{2}}^{(n)^c}\right\} < \frac{\epsilon}{2}$, $\forall n \ge N_2$.
Finally, by taking $N_0 = \max\{N_1, N_2\}$, we can conclude that $\mathsf{Pr}\left\{\hat{S}_n = S_n\right\} \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$, $\forall n \ge N_0$.
Then the claim is proved.
Therefore, for any $\epsilon \in (0,1)$, once n is large enough, any (n,k) code has error probability greater than $1-\epsilon$. So, we can conclude that $\lim_{n\to\infty} \epsilon^*(n,k) = 1$.

Grading Policy

(10) Flawless proof
(8-6) Reasonable procedure but with some missing detail, e.g. undetermined δ
(4) Conceptually wrong, e.g. bound the error probability for only typicality-based code
(2) Wrong but I can tell you have tried

2. (Entropy calculation) [16]

a) Let $X_1, X_2, \ldots, X_n$ be n discrete random variables with disjoint alphabets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ respectively. Let J be a random index, independent of everything else, taking values in $\{1,2,\ldots,n\}$ with

$$\mathsf{Pr}\{J = j\} = p_j,\ \ j = 1,2,\ldots,n.$$

Find $\mathrm{H}(X_J)$ in terms of $\mathrm{H}(X_1), \mathrm{H}(X_2), \ldots, \mathrm{H}(X_n)$ and $p_1, p_2, \ldots, p_n$. [8]

b) A biased coin (Head with probability p and Tail with probability $(1-p)$) is flipped until the first Head occurs. Let N denote the number of flips required.

Find $\mathrm{H}(N)$ if it exists, or show that it does not exist.

Solution:

2. a)

Note that $\mathrm{H}(J|X_J) = 0$ since the alphabets are disjoint. Hence, $\mathrm{H}(X_J) = \mathrm{H}(X_J, J) - \mathrm{H}(J|X_J) = \mathrm{H}(X_J, J)$.
Then,

$$\begin{aligned} \mathrm{H}(X_J, J) &= \mathrm{H}(J) + \mathrm{H}(X_J|J) && \text{(by chain rule)} \\ &= \sum_{j=1}^{n} -p_j \log p_j + \sum_{j=1}^{n} p_j \mathrm{H}(X_J|J=j) && \text{(by definition)} \end{aligned}$$

$$= \sum_{j=1}^{n} -p_j \log p_j + \sum_{j=1}^{n} p_j \mathrm{H}(X_j)$$

Grading Policy

(8) Flawless proof and correct answer
(6) Correct answer but missing detail about the use of chain rule
(4-0) Wrong answer

2. b)

Note that $\mathsf{p}_N(n) := \mathsf{Pr}\{N = n\} = (1-p)^{n-1}p$.
Thus, by definition, $\mathrm{H}(N) = \sum_{n=1}^{\infty} -\mathsf{p}_N(n) \log \mathsf{p}_N(n)$ if it exists.

$$\begin{aligned}
\sum_{n=1}^{\infty} -\mathsf{p}_N(n) \log \mathsf{p}_N(n) &= \sum_{n=1}^{\infty} -(1-p)^{n-1} p \log\left((1-p)^{n-1}p\right) \\
&= -p\log(1-p) \sum_{n'=0}^{\infty} n'(1-p)^{n'} - p\log p \sum_{n'=0}^{\infty} (1-p)^{n'} \\
&= -p\log(1-p) \cdot \frac{1-p}{p^2} - p\log p \cdot \frac{1}{p} \\
&= \frac{-(1-p)\log(1-p) - p\log p}{p} \qquad \text{, exist if } p > 0
\end{aligned}$$

The third equality comes from the following. Let $r < 1$ and $S_n := \sum_{i=0}^{n} i r^i = 1 \cdot r + 2 \cdot r^2 + \ldots + n \cdot r^n$.
Thus, $rS_n = 1 \cdot r^2 + 2 \cdot r^3 + \ldots + n \cdot r^{n+1}$.
Then we have $(1-r)S_n = r + r^2 + \ldots + r^n - n \cdot r^{n+1} = \frac{r(1-r^n)}{1-r} - n \cdot r^{n+1}$.
That results in $S_n = \frac{r(1-r^n)}{(1-r)^2} - \frac{n \cdot r^{n+1}}{1-r} \longrightarrow \frac{r}{(1-r)^2}$ as $n \to \infty$.

Grading Policy

(8) Flawless proof and correct answer
(7) Almost correct but ignore the case when $p = 0$
(6-0) More mistakes in the derivation

3. (Mixing Increases Entropy) [14]

Consider a probability vector $\boldsymbol{p} = (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_d)$.

a) For $i \neq j \in \{1, 2, \ldots, d\}$, an $\{i, j\}$-mixing of $\boldsymbol{p}$, called $\boldsymbol{p}_{\{i,j\}}$, is another probability vector where both the i-th and the j-th coordinates are replaced by $\frac{p_i + p_j}{2}$.

Show that

$$\mathrm{H}(\boldsymbol{p}) \leq \mathrm{H}\left(\boldsymbol{p}_{\{i,j\}}\right). \qquad [8]$$

b) For $\mathcal{I} \subseteq \{1, 2, \ldots, d\}$, an $\mathcal{I}$-mixing of $\boldsymbol{p}$, $\boldsymbol{p}_{\mathcal{I}}$, is another probability vector where for all the $i \in \mathcal{I}$, the i-th coordinate p_i is replaced by $\frac{\sum_{i \in \mathcal{I}} p_i}{|\mathcal{I}|}$.

Show that

$$\mathrm{H}(\boldsymbol{p}) \leq \mathrm{H}(\boldsymbol{p}_{\mathcal{I}}). \qquad [6]$$

Solution:

3. a) b)

We claim that: The function $f : \mathbb{R}_+ \to \mathbb{R}$ defined by $f(p) := \begin{cases} 0 & \text{, if } p = 0 \\ -p \log p & \text{, if } p > 0 \end{cases}$ is concave in $[0, 1]$.

The claim can be easily shown by showing that of second derivative of $f(\cdot)$ in $(0, 1]$ is negative and that $f((1 - \alpha) \cdot 0 + \alpha p) \geq (1 - \alpha) f(0) + \alpha f(p)$.

In this case, for any $\mathcal{I} \subseteq \{1, 2, ..., d\}$,

$$\begin{aligned}
\mathrm{H}(\boldsymbol{p}_{\mathcal{I}}) &:= \sum_{i \in \{1,...,d\} \setminus \mathcal{I}} -p_i \log p_i + \sum_{i \in \mathcal{I}} -\bar{p} \log \bar{p} \qquad \text{, where } \bar{p} := \frac{\sum_{j \in \mathcal{I}} p_j}{|\mathcal{I}|} \\
&= \sum_{i \in \{1,...,d\} \setminus \mathcal{I}} -p_i \log p_i + |\mathcal{I}| \, f\Big(\frac{\sum_{j \in \mathcal{I}} p_j}{|\mathcal{I}|}\Big) \\
&\geq \sum_{i \in \{1,...,d\} \setminus \mathcal{I}} -p_i \log p_i + |\mathcal{I}| \sum_{j \in \mathcal{I}} \frac{1}{|\mathcal{I}|} f(p_j) \\
&= \mathrm{H}(\boldsymbol{p})
\end{aligned}$$

And clearly, b) implies a).

Grading Policy

(14) Flawless proof
(13-12) Almost correct but fail to cover the case when the entry is 0
(10) Reasonable procedure but with unclear derivation and abuse of notation
(8-0) More mistakes in derivation