ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 35 — Information Theory and Coding

Solutions to Final exam — Jan. 13, 2025

PROBLEM 1. (9 points)

(a) (2 pts) Suppose X and Y are random variables and suppose that X is uniformly distributed on a finite set of values. Let $p = \Pr(X = Y)$. Show that $I(X;Y) \geq pH(X) - 1$.
Hint: Upper bound $H(X|Y)$ by Fano's inequality. (Note that in this exercise Y need not be discrete.)

Solution: By Fano's inequality, we have $H(X|Y) \leq h_2(p) + (1-p)\log(|\mathcal{X}|-1)$. Note that $h_2(p) \leq 1$ and $\log(|\mathcal{X}|-1) \leq \log|\mathcal{X}| = H(X)$ since X is uniform, and we are done.

Consider now a discrete memoryless channel for which the input $x \in [0,1]$ and the output $Y \in [0,1]$ are related via $Y = \min(x, Z)$ where Z is a random variable uniformly distributed on $[0,1]$.

(b) (2 pts) Fix a positive integer k and a positive number c with $0 < c \leq 1$. Suppose the input X is uniformly distributed on $\left\{\frac{ci}{k} : i = 0, \dots, k\right\}$. Show that $I(X;Y) \geq (1-c)\log(1+k) - 1$.

Solution: Note that X is uniformly distributed on $k+1$ values, hence $H(X) = \log(k+1)$. Also note that for $p = \Pr(X = Y) = \sum_{i=0}^{k} \Pr(Z \geq \frac{ci}{k}) \geq \Pr(Z \geq c) = 1-c$. Plugging $p = 1-c$ and $H(X) = \log(k+1)$ in the inequality shown in (a), we are done.

(c) (2 pts) Find the capacity of this channel.

Solution: For any positive integer k and $c < 1$, we know from (b) that $I(X;Y) \geq (1-c)\log(1+k)-1$. Hence, the capacity is infinity. (Suppose it were finite, say equal to C, then there exists a sufficiently large k such that $(1-c)\log(1+k)-1 > C \geq I(X;Y)$, which is a contradiction.)

(d) (3 pts) Fix $0 \leq a < b \leq 1$. Suppose now the input x is constrained to be in the interval $[a,b]$. Find the capacity of the channel under this constraint.
Hint: Pick $c < b-a$, and let X be uniformly distributed on $\left\{a + \frac{ci}{k} : i = 0, \dots, k\right\}$.

Solution: As given in the hint, let $c < b - a$, and let X be uniformly distributed on $\left\{a + \frac{ci}{k} : i = 0, \dots, k\right\}$. Again, $H(X) = \log(k+1)$ and $p \geq \Pr(Z \geq a + c) = 1 - a - c > 0$. Thus, the capacity is infinity once again, for any choice of a, b such that $a < b$.

Remarks: It is surprising that the capacity turns out to be infinity in all cases, including when a and b are arbitrarily close to each other, as long as they are different (if $a = b$, the capacity is trivially zero since $I(X;Y) \leq \log|\mathcal{X}| = 0$). It is also surprising that the capacity does not depend on the position of a, b within $[0,1]$.

Problem 2. (7 points)

Recall that the Hamming weight $w_H(x^n)$ of a binary vector x^n is the number of 1's that occur in x^n, i.e., $w_H(x^n) = \sum_{i=1}^n \mathbb{1}(x_i = 1)$. Suppose $X^n \in \{0,1\}^n$ is a random binary vector, with $\frac{1}{n}\mathbb{E}[w_H(X^n)] = p$.

(a) (1 pts) Let $p_i = \Pr(X_i = 1)$. How are $p_1, \dots, p_n$ and p related?

Solution: p is the arithmetic average of p_i's, as

$$p = \frac{1}{n}\mathbb{E}[w_H(X^n)] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \mathbb{1}(x_i = 1)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathbb{1}(x_i = 1)\right]$$
$$= \frac{1}{n}\sum_{i=1}^n p_i.$$

(b) (1 pts) How are $H(X_i)$ and p_i related?

Solution: Since X_i is a binary random variable taking the values 0 and 1 with probability $1-p$ and p, by definition, $H(X_i) = h_2(p_i)$.

(c) (3 pts) Show that $\frac{1}{n}H(X^n) \le h_2(p)$.

Solution: This follows from the following chain of inequalities:

$$\frac{1}{n}H(X^n) \le \frac{1}{n}\sum_{i=1}^n H(X_i) = \frac{1}{n}\sum_{i=1}^n h_2(p_i)$$
$$\le h_2\left(\frac{1}{n}\sum_{i=1}^n p_i\right) = h_2(p),$$

which follow from the subadditivity of entropy and the concavity of h_2.

(d) (2 pts) Let $B_n(r)$ be the Hamming ball of radius r around 0^n, i.e., $B_n(r) = \{x^n \in \{0,1\}^n : w_H(x^n) \le r\}$. For $r \le \frac{n}{2}$, show that $\frac{1}{n}\log|B_n(r)| \le h_2(\frac{r}{n})$.

Hint: Let X^n be uniformly distributed on $B_n(r)$.

Solution: As given in the hint, let X^n be uniformly distributed on $B_n(r)$. Then, $H(X^n) = \log|B_n(r)|$. For any $x^n \in B_n(r)$, we have $w_H(x^n) \le r$, hence $p = \frac{1}{n}\mathbb{E}[w_H(X^n)] \le \frac{r}{n} \le \frac{1}{2}$. Observing that $p \mapsto h_2(p)$ is increasing for $p \in [0, \frac{1}{2}]$, we are done by (c).

Remarks: This strengthens one of the homework exercises that the Hamming *sphere* $S_n(r) = \{x^n : w_H(x^n) = r\}$ has volume less than $2^{nh_2(r/n)}$, by saying that not only the surface of the sphere, but the entire volume of vectors of weight at most r is upper bounded by $2^{nh_2(r/n)}$.

Problem 3. (7 points)

Suppose X is an *integer valued* random variable, Z is uniformly distributed on the interval $[0,1]$ and is independent of X, and $Y = X + Z$.

(a) (2 pts) How is $H(X)$ and $h(Y)$ related?

Solution: $H(X) = h(Y)$, this can be seen by first observing that $f_Y(y) = \Pr(X = \lfloor y \rfloor)$, so $\int f_Y(y)\log f_Y(y)dy = \sum_k \Pr(X=k)\log\Pr(X=k)$.

(b) (2 pts) Show that $\mathrm{Var}(Y) = \mathrm{Var}(X) + \frac{1}{12}$.

Solution: Since X and Z are independent, $\mathrm{Var}(Y) = \mathrm{Var}(X)+\mathrm{Var}(Z) = \mathrm{Var}(X)+\frac{1}{12}$, as $\mathrm{Var}(Z) = \int_0^1 z^2\, dz - \left(\int_0^1 z\, dz\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

(c) (1 pts) Show that $H(X) \le \frac{1}{2}\log\left(2\pi e\left(\mathrm{Var}(X) + \frac{1}{12}\right)\right)$.

Solution: For any continuous random variable Y, we have $h(Y) \le (1/2)\log[2\pi e \mathrm{Var}(Y)]$. Hence, the desired inequality follows from (b).

(d) (2 pts) Suppose $S_n = \sum_{i=1}^n B_i$ where B_i are i.i.d. Bern($\frac{1}{2}$). Show that $H(S_n) \le \frac{1}{2}\log\left(2\pi e(\frac{n}{4} + \frac{1}{12})\right) \le \frac{1}{2}\log(n+1) + \frac{1}{2}\log\left(\frac{\pi e}{2}\right)$

Solution: Let $X = S_n$ and let Y and Z be as described in the parts above. Then $\mathrm{Var}(X) = \mathrm{Var}(S_n) = n\mathrm{Var}(B_i) = \frac{n}{4}$. Hence, by (c), we have

$$\begin{aligned} H(S_n) &\le \frac{1}{2}\log\left(2\pi e\left(\mathrm{Var}(S_n) + \frac{1}{12}\right)\right) = \frac{1}{2}\log\left(2\pi e\left(\frac{n}{4} + \frac{1}{12}\right)\right) \\ &= \frac{1}{2}\log\left(\frac{\pi e}{2}\right) + \frac{1}{2}\log\left(n + \frac{1}{3}\right), \end{aligned}$$

and we are done as $\log(n + \frac{1}{3}) \le \log(n+1)$.

Remarks: The final result follows immediately from (c), but it captures the "correct" order of increase of $H(S_n)$, i.e., $\lim_{n\to\infty} \frac{H(S_n)}{\log n} = \frac{1}{2}$.

Problem 4. (11 points)

We are given a binary input channel $W : \mathbb{F}_2 \to \mathcal{Y}$. Let $Q(W) = \sum_y \sqrt{W(y|0)W(y|1)}$.

(a) (1 pts) Find $Q(\mathrm{BEC}(p))$.

Solution: By simply computing the expression, we get

$$\begin{aligned} Q(\mathrm{BEC}(p)) &= \sum_y \sqrt{\mathrm{BEC}(y|0)\mathrm{BEC}(y|1)} \\ &= \sqrt{(1-p)\cdot 0} + \sqrt{p \cdot p} + \sqrt{0 \cdot (1-p)} = p. \end{aligned}$$

(b) (3 pts) Suppose the channel input X is equally likely to be 0 or 1, and upon observing the channel output $Y = y$, we estimate the value of X as $\hat{x}(y) = 0$ if $W(y|0) > W(y|1)$; 1 else. Show that $\Pr(\hat{x}(Y) \ne X) \le Q(W)$.

Hint: First condition on $X = 0$. In this case $\mathbb{1}\{\hat{x}(y) \ne X\} \le \sqrt{\frac{W(y|1)}{W(y|0)}}$.

Solution: We can write the error probability as

$$\Pr(\hat{x}(Y) \ne X) = \Pr(X = 0)\Pr(\hat{x}(Y) \ne 0 \mid X = 0) + \Pr(X = 1)\Pr(\hat{x}(Y) \ne 1 \mid X = 1).$$

Using the hint, note that $\mathbb{1}\{\hat{x}(y) \ne 0\} \le \sqrt{\frac{W(y|1)}{W(y|0)}}$ and hence,

$$\begin{aligned} \Pr(\hat{x}(Y) \ne 0 \mid X = 0) &\le \sum_y W(y|0)\mathbb{1}\{\hat{x}(y) \ne 0\} \\ &\le \sum_y W(y|0)\sqrt{\frac{W(y|1)}{W(y|0)}} = \sum_y \sqrt{W(y|0)W(y|1)}. \end{aligned}$$

Similarly, $\Pr(\hat{x}(Y) \neq 0 \mid X = 0) \leq \sum_y \sqrt{W(y|0)W(y|1)}$. Hence, the average $\Pr(\hat{x}(Y) \neq X)$ is also smaller than $\sum_y \sqrt{W(y|0)W(y|1)} = Q(W)$.

Recall the polar construction which, from two instances of the channel W synthesized the channels $W^- : \mathbb{F}_2 \to \mathcal{Y}^2$ and $W^+ : \mathbb{F}_2 \to \mathcal{Y}^2 \times \mathbb{F}_2$ with

$$W^-(y_1y_2|u_1) = \frac{W(y_1|u_1)W(y_2|0) + W(y_1|u_1 \oplus 1)W(y_2|1)}{2}$$

and

$$W^+(y_1y_2u_1|u_2) = \frac{1}{2}W(y_1|u_1 \oplus u_2)W(y_2|u_2).$$

(c) (2 pts) Show that $Q(W^+) = Q(W)^2$.
Hint: $Q(W^+) = \sum_{y_1,y_2,u_1} \sqrt{W^+(y_1y_2u_1|0)W^+(y_1y_2u_1|1)}$

Solution: By simplifying the expression, we have

$$\begin{aligned}
Q(W^+) &= \sum_{y_1,y_2,u_1} \sqrt{W^+(y_1y_2u_1|0)W^+(y_1y_2u_1|1)} \\
&= \sum_{y_1,y_2,u_1} \sqrt{\frac{1}{2}W(y_1|u_1 \oplus 0)W(y_2|0)\frac{1}{2}W(y_1|u_1 \oplus 1)W(y_2|1)} \\
&= \frac{1}{2} \sum_{y_1,y_2,u_1} \sqrt{W(y_1|u_1 \oplus 0)W(y_1|u_1 \oplus 1)W(y_2|0)W(y_2|1)} \\
&= \frac{1}{2} \sum_{y_1,u_1} \sqrt{W(y_1|u_1 \oplus 0)W(y_1|u_1 \oplus 1)} \sum_{y_2} \sqrt{W(y_2|0)W(y_2|1)} \\
&= \frac{1}{2} \cdot 2Q(W) \cdot Q(W) = Q(W)^2.
\end{aligned}$$

(d) (2 pts) Use the inequality

$$\sqrt{(ab+cd)(ac+bd)} \leq \left(\sqrt{ab} + \sqrt{cd}\right)\left(\sqrt{ac} + \sqrt{bd}\right) - 2\sqrt{abcd}$$

to show that $Q(W^-) \leq 2Q(W) - Q(W)^2$.

Solution: Start by writing $2Q(W^-)$ as a double sum over y_1, y_2, and note that the terms are exactly of the form ($\sqrt{(ab+cd)(ac+bd)}$ with $a = W(y_1|0)$, $b = W(y_2|0)$, $c = W(y_1|1)$, $d = W(y_2|1)$. The given inequality then gives us exactly what we want. From the "$-2\sqrt{abcd}$" term we get $-2Q(W)^2$; each of the four terms in the expansion of $(\sqrt{ab} + \sqrt{cd})(\sqrt{ac} + \sqrt{bd})$ gives $Q(W)$.

(e) (3 pts) Given a binary input channel W, Let $\tilde{W} = \mathrm{BEC}(p)$, where $p = Q(W)$. Show that for any sign sequence $s^t \in \{+,-\}^t$, $Q(W^{s^t}) \leq Q(\tilde{W}^{s^t})$.

Solution: Let $Q_i = Q(W^{s^i})$ and similarly $\tilde{Q}_i = Q(\tilde{W}^{s^i})$. $Q_t \leq \tilde{Q}_t$ follows by induction on t. The base case ($t = 0$) holds trivially. Since the Successive terms of the $\tilde{Q}_i$ sequence are found (with equality, since this is a BEC) from the previous term by $x \mapsto x^2$ or $x \mapsto 2x - x^2$ operations depending on the sign s_i. Note that both these operations are monotonically increasing. Meanwhile, the Q_i sequence goes through corresponding operations but with inequality (because of (d)), and we are done.

Remarks: The inequality in part (b) is the Bhattacharyya bound on the error probability. An immediate consequence of parts (b) and (e) is the following: With W and $\tilde{W}$ as above, suppose enc is a polar code designed for the channel $\tilde{W}$. Let $\tilde{p}_e$ denote the error probability of the code enc (with the corresponding polar decoder) when used on $\tilde{W}$, and let p_e denote the error probability of of the same code when used on channel W (with the corresponding polar decoder). Then $p_e \leq \tilde{p}_e$.