Sample problems for the 13th lecture (June 14)

1. (a) Find the average number of random bits executed per output symbol in the below program.

```
For i = 1 to i = n do the following
{
  Flip-a-fair-coin;                    \\ one random bit
  If "Head", then output 0;
  else
       {
       Flip-a-fair-coin;               \\ one random bit
       If "Head", then output −1;
       else output 1;
       }
}
```

   (b) Find the probabilities of the random output sequence X_1, X_2, ..., X_n of the above algorithm. What is the entropy rate of this random output sequence? Is the above algorithm asymptotically optimal in the sense of minimizing the average number of random bits executed per output symbol among all algorithms that generate the random outputs of the same statistics?

   (c) What is the resolution rate of X_1, X_2, ..., X_n in (b)?

   (d) Repeat (a), (b) and (c) for the algorithm below.

```
        i = 0
        while (1)
        {
        even = False;
        while (1)
          {Flip-a-fair-coin;       \\ one random bit
          if (Head)
            {if (even==True) { output 0; goto next;}
             else {output 1; goto next;}
            }
          else
            {if (even==True) even=False;
             else even=True;
            }
          }
next:   i = i + 1; if i > n, then break;
        }
```

Solution.

(a) On an average, 1.5 random bits are executed per output symbol.

(b) For each i, $P_{X_i}(-1) = 1/4$, $P_{X_i}(0) = 1/2$, and $P_{X_i}(1) = 1/4$, and $X_1, X_2, \ldots, X_n$ are i.i.d. Its entropy rate is 1.5 bits. Since entropy rate is equal to the average number of random bits executed per output symbol, the algorithm is asymptotically optimal.

(c) Each X_i is 4-type. Hence, its resolution rate is $\frac{1}{n}R(X^n) = \log_2(4) = 2$ bits.

(d) The average complexity per output symbol of this algorithm is

$$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \cdots = 2 \text{ random bits.}$$

The distributions of each X_i are

$$P_{X_i}(1) = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \cdots = \frac{2}{3}$$

and

$$P_{X_i}(0) = 1 - P_{X_i}(1),$$

and $X_1, X_2, \ldots, X_n$ are i.i.d. Its entropy rate is $\frac{2}{3}\log_2(\frac{3}{2}) + \frac{1}{3}\log_2(3) = \log_2(3) - \frac{2}{3} \approx 0.918$.

The resolution rate is $\log_2(3) \approx 1.58$.

The algorithm is not asymptotically optimal since the average number of random bits executed per output symbol is strictly larger than the entropy rate.

2. (a) Assume the cardinality of X is finite. Show that

$$H(X) \leq R(X),$$

where $H(X)$ is the entropy of X, and $R(X)$ is the resolution of X. When will equality holds for this inequality?

Hint: $R(X) = \log_2(M)$ if X has an M-type probability distribution, and $H(X) = \sum_{x \in \mathcal{X}} P_X(x) \log_2 \frac{1}{P_X(x)}$.

(b) Between the two constraints

$$\frac{1}{n}H(\tilde{X}^n) < R + \gamma \quad \text{and} \quad \frac{1}{n}R(\tilde{X}^n) < R + \gamma,$$

which one is used to define the ϵ-mean-achievable resolution rate, and which one is used to define the ϵ-achievable resolution rate? Which constraint is stronger?

Solution.

(a) First, we note that $|\mathcal{X}| \le M$ because

$$1 = \sum_{x\in\mathcal{X}} P_X(x) \ge \sum_{x\in\mathcal{X}} \frac{1}{M} = \frac{|\mathcal{X}|}{M}.$$

Then, we have

$$H(X) \le \log_2(|\mathcal{X}|) \le \log_2(M) = R(X).$$

The first inequality holds with equality when X is uniform; the second inequality holds with equality when $|\mathcal{X}| = M$. As a summary, $H(X) = R(X)$ iff $P_X(x) = \frac{1}{M}$ for every $x \in \mathcal{X}$.

(b) $\frac{1}{n}H(\tilde{X}^n) < R + \gamma$ is used to defined the ϵ-mean-achievable resolution rate, and $\frac{1}{n}R(\tilde{X}^n) < R + \gamma$ is used to define ϵ-achievable resolution rate.
Since $\frac{1}{n}H(\tilde{X}^n) \le \frac{1}{n}R(\tilde{X}^n)$, $\frac{1}{n}R(\tilde{X}^n) < R + \gamma$ is a stronger (more restricted) contraint than $\frac{1}{n}H(\tilde{X}^n) < R + \gamma$.

3. Suppose we transmit uniform inputs X_1, X_2 over the 2-fold BSC with crossover probability ϵ, where $0 < \epsilon < \frac{1}{2}$.

(a) List the normalized information densities for all input-output pairs.

(b) Plot the cdf of the normalized information density for $\epsilon = \frac{1}{4}$.

(c) Define a "typical" Feinstein's set as

$$\mathcal{G} := \left\{ (x^2, y^2) \in \mathcal{X}^2 \times \mathcal{Y}^2 \ : \ \frac{1}{2} i_{X^2W^2}(x^2; y^2) \ge \frac{1}{2}\log_2(M) + \gamma \right\}.$$

Specify $\mathcal{G}$ for $M = 2$ and $\gamma = 0.01$ under $\epsilon = \frac{1}{4}$.

Solution.

(a)

$\frac{1}{2}\log_2 \frac{P_{Y^2\|X^2}(y^2\|x^2)}{P_{Y^2}(y^2)}$	$y_1y_2 = 00$	$y_1y_2 = 01$	$y_1y_2 = 10$	$y_1y_2 = 11$
$x_1x_2 = 00$	$\frac{1}{2}\log_2[4(1-\epsilon)^2]$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4\epsilon^2]$
$x_1x_2 = 01$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4(1-\epsilon)^2]$	$\frac{1}{2}\log_2[4\epsilon^2]$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$
$x_1x_2 = 10$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4\epsilon^2]$	$\frac{1}{2}\log_2[4(1-\epsilon)^2]$	$\frac{1}{2}\log_2[4\epsilon^2]$
$x_1x_2 = 11$	$\frac{1}{2}\log_2[4\epsilon^2]$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4\epsilon(1-\epsilon)]$	$\frac{1}{2}\log_2[4(1-\epsilon)^2]$

(b) First,

$$\frac{1}{2}\log_2\frac{P_{Y^2|X^2}(Y^2|X^2)}{P_{Y^2}(Y^2)} = \frac{1}{2}\log_2[4(1-\epsilon)^2] = \frac{1}{2}\log_2(\frac{9}{4}) \approx 0.585$$

occurs with probability

$$\begin{aligned} & P_{X^2}(00)P_{Y^2|X^2}(00|00) + P_{X^2}(01)P_{Y^2|X^2}(01|01) \\ & +P_{X^2}(10)P_{Y^2|X^2}(10|10) + P_{X^2}(11)P_{Y^2|X^2}(11|11) \\ = \;& 4 \times \frac{1}{4} \times (1-\epsilon)^2 = (1-\epsilon)^2 = \frac{9}{16}. \end{aligned}$$

Next,

$$\frac{1}{2}\log_2\frac{P_{Y^2|X^2}(Y^2|X^2)}{P_{Y^2}(Y^2)} = \frac{1}{2}\log_2[4\epsilon^2] = -1$$

occurs with probability

$$\begin{aligned} & P_{X^2}(00)P_{Y^2|X^2}(11|00) + P_{X^2}(01)P_{Y^2|X^2}(10|01) \\ & +P_{X^2}(10)P_{Y^2|X^2}(01|10) + P_{X^2}(11)P_{Y^2|X^2}(00|11) \\ = \;& 4 \times \frac{1}{4} \times \epsilon^2 = \epsilon^2 = \frac{1}{16}. \end{aligned}$$

Hence,

$$\frac{1}{2}\log_2\frac{P_{Y^2|X^2}(Y^2|X^2)}{P_{Y^2}(Y^2)} = \frac{1}{2}\log_2[4\epsilon(1-\epsilon)] = \frac{1}{2}\log_2(\frac{3}{4}) \approx -0.208$$

occurs with probability $1-(1-\epsilon)^2-\epsilon^2 = 2\epsilon(1-\epsilon) = \frac{3}{8}$.
(You shall be able to plot the cdf yourself.)

(c) From (b), it is clear that $\mathcal{G} = \{(00,00),(01,01),(10,10),(11,11)\}$.

4. Complete the proof of Feinstein's lemma (indicated in three boxes below).

Lemma 1 (Feinstein's Lemma) Fix a positive n. For every $\gamma > 0$ and input distribution P_{X^n} on $\mathcal{X}^n$, there exists an (n, M) block code for the transition probability $P_{W^n} = P_{Y^n|X^n}$ that its average error probability $P_e(\mathcal{C}_n)$ satisfies

$$P_e(\mathcal{C}_n) < \Pr\left[\frac{1}{n} i_{X^n W^n}(X^n;Y^n) < \frac{1}{n}\log M + \gamma\right] + e^{-n\gamma}.$$

Proof:

Step 1: Notations. Define

$$\mathcal{G}:=\left\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n\ :\ \frac{1}{n}i_{X^nW^n}(x^n;y^n)\geq\frac{1}{n}\log M+\gamma\right\}.$$

Let $\nu:=e^{-n\gamma}+P_{X^nW^n}(\mathcal{G}^c)$.
Feinstein's Lemma obviously holds if $\nu\geq 1$, because then

$$P_e(\mathcal{C}_n)\leq 1\leq\nu:=\Pr\left[\frac{1}{n}i_{X^nW^n}(X^n;Y^n)<\frac{1}{n}\log M+\gamma\right]+e^{-n\gamma}.$$

So we assume $\nu<1$, which immediately results in

$$P_{X^nW^n}(\mathcal{G}^c)<\nu<1,$$

or equivalently,

$$P_{X^nW^n}(\mathcal{G})>1-\nu>0. \tag{1}$$

Therefore, denoting

$$\mathcal{A}:=\{x^n\in\mathcal{X}^n\ :\ P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n)>1-\nu\}$$

with $\mathcal{G}_{x^n}:=\{y^n\in\mathcal{Y}^n\ :\ (x^n,y^n)\in\mathcal{G}\}$, we have

$$P_{X^n}(\mathcal{A})>0,$$

because if $P_{X^n}(\mathcal{A})=0$,

$$(\forall\ x^n \text{ with } P_{X^n}(x^n)>0)\ P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n)\leq 1-\nu$$

$$\Rightarrow\sum_{x^n\in\mathcal{X}^n}P_{X^n}(x^n)P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n)=P_{X^nW^n}(\mathcal{G})\leq 1-\nu,$$

and a contradiction to (1) is obtained.

Step 2: Encoder. Choose an x_1^n in $\mathcal{A}$ (Recall that $P_{X^n}(\mathcal{A})>0$.) Define $\Gamma_1=\mathcal{G}_{x_1^n}$. (Then $P_{Y^n|X^n}(\Gamma_1|x_1^n)>1-\nu$.)
Next choose, if possible, a point $x_2^n\in\mathcal{X}^n$ without replacement (i.e., x_2^n cannot be identical to x_1^n) for which

$$P_{Y^n|X^n}\left(\mathcal{G}_{x_2^n}-\Gamma_1\middle|\,x_2^n\right)>1-\nu,$$

and define $\Gamma_2:=\mathcal{G}_{x_2^n}-\Gamma_1$.

Continue in the following way as for codeword i: choose x_i^n to satisfy

$$P_{Y^n|X^n}\left(\mathcal{G}_{x_i^n} - \bigcup_{j=1}^{i-1}\Gamma_j \middle| x_i^n\right) > 1-\nu,$$

and define $\Gamma_i := \mathcal{G}_{x_i^n} - \bigcup_{j=1}^{i-1}\Gamma_j$.

Repeat the above codeword selecting procedure until either M codewords are selected or all the points in $\mathcal{A}$ are exhausted.

Step 3: Decoder. Define the decoding rule as

$$\phi(y^n) = \begin{cases} i, & \text{if } y^n \in \Gamma_i \\ \text{arbitrary}, & \text{otherwise.} \end{cases}$$

Step 4: Probability of error. For all selected codewords, the error probability given codeword i is transmitted, $\lambda_{e|i}$, satisfies

$$\lambda_{e|i} \le P_{Y^n|X^n}(\Gamma_i^c|x_i^n) < \nu.$$

(Note that $(\forall\ i)\ P_{X^n|X^n}(\Gamma_i|x_i^n) \ge 1-\nu$ by Step 2.) Therefore, if we can show that the above codeword selecting procedures will not terminate before M, then

$$P_e(\mathcal{C}_n) = \frac{1}{M}\sum_{i=1}^{M}\lambda_{e|i} < \nu.$$

Step 5: Claim. The codeword selecting procedure in Step 2 will not terminate before M.

Proof: We will prove it by contradiction.

Suppose the above procedure terminates before M, say at $N < M$. Define the set

$$\mathcal{F} := \bigcup_{i=1}^{N}\Gamma_i \in \mathcal{Y}^n.$$

Consider the probability

$$P_{X^nW^n}(\mathcal{G}) = P_{X^nW^n}[\mathcal{G}\cap(\mathcal{X}^n\times\mathcal{F})] + P_{X^nW^n}[\mathcal{G}\cap(\mathcal{X}^n\times\mathcal{F}^c)]. \quad (2)$$

Since for any $y^n \in \mathcal{G}_{x_i^n}$,

$$P_{Y^n}(y^n) \le \frac{P_{Y^n|X^n}(y^n|x_i^n)}{M\cdot e^{n\gamma}},$$

we have

$$\begin{aligned} P_{Y^n}(\Gamma_i) &\leq P_{Y^n}(\mathcal{G}_{x_i^n}) \\ &\leq \frac{1}{M} e^{-n\gamma} P_{Y^n|X^n}(\mathcal{G}_{x_i^n}) \\ &\leq \frac{1}{M} e^{-n\gamma}. \end{aligned} \tag{3}$$

(a) Show that

$$P_{X^nW^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F})] \leq \frac{N}{M} e^{-n\gamma}.$$

Hint: Use (3).

(b) Show that

$$P_{X^nW^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F}^c)] \leq 1 - \nu.$$

Hint: For all $x^n \in \mathcal{X}^n$,

$$P_{Y^n|X^n}\left(\mathcal{G}_{x^n} - \bigcup_{i=1}^{N} \Gamma_i \middle| x^n\right) \leq 1 - \nu.$$

(c) Prove that (a) and (b) joinly imply $N \geq M$, resulting in a contradiction.
Hint: By definition of $\mathcal{G}$, $P_{X^nW^n}(\mathcal{G}) = 1 - \nu + e^{-n\gamma}$.

Solution. See the note or slides.