Information Theory: Final Exam on 21 June 2019

1. (a) (6%) What are the three axioms raised by Shannon for the measurement of information?

   (b) (6%) Define the *divergence typical set* as

$$\mathcal{A}_n(\delta) := \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n} \log_2 \frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} - D(P_X \| P_{\hat{X}}) \right| < \delta \right\}.$$

   It can be shown that for any sequence x^n in $\mathcal{A}_n(\delta)$,

$$P_{X^n}(x^n) 2^{-n(D(P_X\|P_{\hat{X}})-\delta)} > P_{\hat{X}^n}(x^n) > P_{X^n}(x^n) 2^{-n(D(P_X\|P_{\hat{X}})+\delta)}.$$

   Prove that

$$P_{\hat{X}^n}(\mathcal{A}_n(\delta)) \le 2^{-n(D(P_X\|P_{\hat{X}})-\delta)} P_{X^n}(\mathcal{A}_n(\delta)).$$

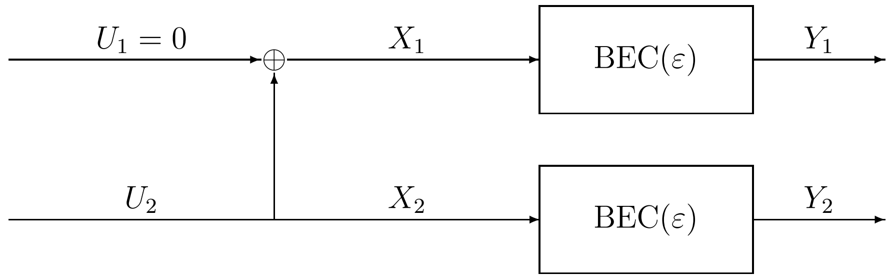
Solution.

(a) i) Monotonicity in event probability
    ii) Additivity for independent events
    iii) Continuity in event probability

(b)

$$\begin{aligned}
P_{\hat{X}^n}(\mathcal{A}_n(\delta)) &= \sum_{x^n \in \mathcal{A}_n(\delta)} P_{\hat{X}^n}(x^n) \\
&\le \sum_{x^n \in \mathcal{A}_n(\delta)} P_{X^n}(x^n) 2^{-n(D(P_X\|P_{\hat{X}})-\delta)} \\
&= 2^{-n(D(P_X\|P_{\hat{X}})-\delta)} \sum_{x^n \in \mathcal{A}_n(\delta)} P_{X^n}(x^n) \\
&= 2^{-n(D(P_X\|P_{\hat{X}})-\delta)} P_{X^n}(\mathcal{A}_n(\delta))
\end{aligned}$$

2. (6%) For the basic transformation of polar code shown below, answer the following questions.



Note that here we assume the two channels are independent binary erasure channels. Determine the erasure probability for U_2, given $U_1 = 0$ (i.e., frozen as zero)?

Solution. From

$$\mathbb{Q}^+ : \quad U_2 = \begin{cases} Y_1 \oplus U_1 = Y_1, & \text{if } Y_1 \in \{0,1\} \\ Y_2, & \text{if } Y_2 \in \{0,1\} \\ ?\,, & \text{if } Y_1 = Y_2 = E \end{cases}$$

U_2 is erased iff both Y_1 and Y_2 equal erasure, which occurs with probability ε^2.

3. (a) (6%) Let X be a random variable with pdf $f_X(x)$ and support $\mathcal{X}$. Prove that for any function $q(\cdot)$ and positive C,

$$C \cdot E[q(X)] - h(X) \geq \ln(D)$$

where $h(X)$ is the differential entropy of X and

$$D = \frac{1}{\int_{\mathcal{X}} e^{-C\cdot q(X)} dx}.$$

Hint: Reformulate

$$C \cdot E[q(X)] - h(X) = C \cdot \int_{-\infty}^{\infty} f_X(x) q(x) dx - \int_{-\infty}^{\infty} f_X(x) \ln \frac{1}{f_X(x)} dx$$

and use

$$D(X\|Y) = \int_{\mathcal{X}} f_X(x) \ln \frac{f_X(x)}{f_Y(x)} dx \geq 0$$

for any continuous random variable Y that admits a pdf $f_Y(x)$ over support $\mathcal{X}$.

(b) (6%) When does equality hold in (a) such that

$$h(X) = C \cdot E[q(X)] - \ln(D)?$$

(c) (6%) Use (a) and (b) to find the random variable that maximizes the differential entropy among all variables with finite support $[a, b]$.
Hint: Choose $q(x) = 1$ over $[a, b]$ and use (b).

Solution.

(a)

$$\begin{aligned} C \cdot E[q(X)] - h(X) &= C \cdot \int_{-\infty}^{\infty} f_X(x) q(x) dx - \int_{-\infty}^{\infty} f_X(x) \ln \frac{1}{f_X(x)} dx \\ &= \int_{-\infty}^{\infty} f_X(x) \ln \frac{f_X(x)}{e^{-C\cdot q(x)}} dx \\ &= \int_{-\infty}^{\infty} f_X(x) \ln \frac{f_X(x)}{De^{-C_1\cdot q(x)}} dx + \ln(D) \\ &= \int_{-\infty}^{\infty} f_X(x) \ln \frac{f_X(x)}{f_Y(x)} dx + \ln(D) \\ &\geq \ln(D) \end{aligned}$$

where $f_Y(x) = De^{-C\cdot q(x)}$ is a pdf defined over $x \in \mathcal{X}$.

(b) From the derivation in (a), equality holds iff $f_X(x) = f_Y(y)$, i.e., $f_X(x) = De^{-C\cdot q(x)}$ for $x \in \mathcal{X}$.

(c) From (a), we know that

$$h(X) \leq C \cdot E[q(X)] - \ln(D).$$

Thus, the differential entropy is maximized if equality holds in the above inequality. Set $q(x) = 1$. Then, equality holds if

$$f_X(x) = \frac{e^{-C}}{\int_a^b e^{-C} dx} = \frac{1}{b-a}.$$

Consequently, the random variable that maximizes the differential entropy among all variables with finite support $[a, b]$ has a uniform distribution.

4. (8%) Suppose the capacity of k parallel channels, where the ith channel uses power P_i, follows a logarithm law, i.e.,

$$C(P_1, \ldots, P_k) = \sum_{i=1}^{k} \frac{1}{2} \log_2 \left(1 + \frac{P_i}{\sigma_i^2}\right),$$

where σ_i^2 is the noise variance of channel i. Determine the optimal $\{P_i^*\}_{i=1}^k$ that maximize $C(P_1, \ldots, P_k)$ subject to $\sum_{i=1}^k P_i = P$.

Hint: By using the Lagrange multipliers technique and verifying the KKT condition, the maximizer $(P_1, \ldots, P_k)$ of

$$\max \left\{ \sum_{i=1}^{k} \frac{1}{2} \log_2 \left(1 + \frac{P_i}{\sigma_i^2}\right) + \sum_{i=1}^{k} \lambda_i P_i - \nu \left(\sum_{i=1}^{k} P_i - P\right) \right\}$$

can be found by taking the derivative of the above equation (with respect to P_i) and setting it to zero.

Solution. By using the Lagrange multipliers technique and verifying the KKT condition, the maximizer $(P_1, \ldots, P_k)$ of

$$\max \left\{ \sum_{i=1}^{k} \frac{1}{2} \log_2 \left(1 + \frac{P_i}{\sigma_i^2}\right) + \sum_{i=1}^{k} \lambda_i P_i - \nu \left(\sum_{i=1}^{k} P_i - P\right) \right\}$$

can be found by taking the derivative of the above equation (with respect to P_i) and setting it to zero, which yields

$$\lambda_i = \begin{cases} -\dfrac{1}{2\ln(2)} \dfrac{1}{P_i + \sigma_i^2} + \nu = 0, & \text{if } P_i > 0; \\ -\dfrac{1}{2\ln(2)} \dfrac{1}{P_i + \sigma_i^2} + \nu \geq 0, & \text{if } P_i = 0. \end{cases}$$

Hence,

$$\begin{cases} P_i = \theta - \sigma_i^2, & \text{if } P_i > 0; \\ P_i \geq \theta - \sigma_i^2, & \text{if } P_i = 0, \end{cases} \qquad (\text{equivalently, } P_i = \max\{0, \theta - \sigma_i^2\}),$$

where $\theta := \log_2(e)/(2\nu)$ is chosen to satisfy $\sum_{i=1}^{k} P_i = P$.

5. (a) (6%) Prove that the number of x^n's satisfying $P_{X^n}(x^n) \geq \frac{1}{N}$ is at most N.

   (b) (6%) Let $\mathcal{C}_n^*$ be the set that maximizes $\Pr[X^n \in \mathcal{C}_n^*]$ among all sets of the same size M_n. Prove that

$$\Pr[X^n \notin \mathcal{C}_n^*] \leq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n\right],$$

   where $h_{X^n}(X^n) \triangleq \log \frac{1}{P_{X^3}(X^3)}$.

   Hint: Use (a).

   (c) (8%) Let $\mathcal{C}_n$ be a subset of $\mathcal{X}^n$, satisfying that $|\mathcal{C}_n| = M_n$. Prove that for every $\gamma > 0$,

$$\Pr[X^n \notin \mathcal{C}_n] \geq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n + \gamma\right] - \exp\{-n\gamma\}.$$

   Hint: It suffices to prove that

$$\begin{aligned}
\Pr\left[X^n \in \mathcal{C}_n\right] &\leq \Pr\left[\frac{1}{n} h_{X^n}(X^n) \leq \frac{1}{n}\log M_n + \gamma\right] + \exp\{-n\gamma\} \\
&= \Pr\left[\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \frac{1}{n}\log M_n + \gamma\right] + \exp\{-n\gamma\} \\
&= \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n} e^{-n\gamma}\right] + \exp\{-n\gamma\}.
\end{aligned}$$

Solution.

(a) This can be proved by contradiction. Suppose there are $N+1$ x^n's satisfying $P_{X^n}(x^n) \geq \frac{1}{N}$. Then, $1 = \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) \geq \frac{N+1}{N}$, which is a contradiction. Hence, the number of x^n's satisfying $P_{X^n}(x^n) \geq \frac{1}{N}$ is at most N.

(b) From (a), the number of x^n's satisfying $P_{X^n}(x^n) \geq \frac{1}{M_n}$ is at most M_n. Since $\mathcal{C}_n^*$ should consist of M_n words with larger probabilities, we have

$$\begin{aligned}
\Pr[X^n \in \mathcal{C}_n^*] &\geq \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\right] \\
&= \Pr\left[\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq \frac{1}{n}\log M_n\right] \\
&= \Pr\left[\frac{1}{n} h_{X^n}(X^n) \leq \frac{1}{n}\log M_n\right],
\end{aligned}$$

which implies

$$\Pr[X^n \notin \mathcal{C}_n^*] \leq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n\right].$$

(c) We derive

$$
\begin{aligned}
\Pr\left[X^n \in \mathcal{C}_n\right] &= \Pr\left[X^n \in \mathcal{C}_n \text{ and } P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&\quad + \Pr\left[X^n \in \mathcal{C}_n \text{ and } P_{X^n}(X^n) < \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&\leq \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&\quad + \Pr\left[X^n \in \mathcal{C}_n \text{ and } P_{X^n}(X^n) < \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&= \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&\quad + \sum_{x^n \in \mathcal{C}_n} P_{X^n}(x^n) \cdot \mathbf{1}\left\{P_{X^n}(x^n) < \frac{1}{M_n}\,e^{-n\gamma}\right\} \\
&< \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] \\
&\quad + \sum_{x^n \in \mathcal{C}_n} \frac{1}{M_n}\,e^{-n\gamma} \cdot \mathbf{1}\left\{P_{X^n}(x^n) < \frac{1}{M_n}\,e^{-n\gamma}\right\} \\
&= \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] + |\mathcal{C}_n|\frac{1}{M_n}e^{-n\gamma} \\
&= \Pr\left[P_{X^n}(X^n) \geq \frac{1}{M_n}\,e^{-n\gamma}\right] + e^{-n\gamma}.
\end{aligned}
$$

6. (a) (6%) 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) (6%) 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) (6%) What is the resolution rate of X_1, X_2, ..., X_n in (b)?

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, ..., 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.

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

Lemma (**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

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

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

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

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

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

or equivalently,

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

Therefore, denoting

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

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

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

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

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

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

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 $\Gamma_2 \triangleq \mathcal{G}_{x_2^n} - \Gamma_1$.

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

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

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

(Note that $(\forall\ i)\ P_{X^n|X^n}(\Gamma_i|x_i^n) \geq 1 - \nu$ 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

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

Consider the probability

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

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

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

(a) (6%) Show that

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

Hint: Use (3).

(b) (6%) Show that

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

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

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

(c) (6%) Prove that (a) and (b) jointly imply $N \ge M$, 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.