EE276 Homework #7

Due on Friday March 8, 5pm

1. Consider a finite alphabet $\mathcal{X}$. Given $D \geq 0$ and a weight function $\rho : \mathcal{X} \to \mathbb{R}_+$, define

$$B_n(\rho, D) := \left\{ x^n : \frac{1}{n} \sum_{i=1}^{n} \rho(x_i) \leq D \right\}.$$

   (a) Show that

$$B_n(\rho, D) = \bigcup_{p \in \mathbb{P}_n : \langle p, \rho \rangle \leq D} T(p)$$

   where

$$\langle p, \rho \rangle := \sum_{x \in \mathcal{X}} p(x)\rho(x).$$

   (b) Show that

$$|B_n(\rho, D)| \doteq 2^{n \max_{p:\langle p,\rho\rangle \leq D} H(p)}.$$

   (Use the expression derived in (a) to obtain lower and upper bounds on $|B_n(\rho, D)|$ that match up to first order in exponent.)

   (c) Specializing the result of (b), show that for $D \in [0, 1/2]$,

$$\left| \left\{ y^n \in \{0,1\}^n : \frac{1}{n} \sum_{i=1}^{n} y_i \leq D \right\} \right| \doteq 2^{nh_2(D)}$$

   where h_2 is the binary entropy function.

Solution You do not need to write all these details to receive full credit. We write it out in full detail for the sake of the reader.

   (a) Let $p_{x^n} \in \mathbb{P}^n$. Since $\langle \rho, p_{x^n} \rangle = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i)$. Indeed,

$$\begin{aligned}
\langle \rho, p_{x^n} \rangle &= \sum_{x \in \mathcal{X}} \rho(x) p_{x^n}(x) \\
&= \frac{1}{n} \sum_{x \in \mathcal{X}} \rho(x) N(x|x^n) \\
&= \frac{1}{n} \sum_{x \in \mathcal{X}} \rho(x) \sum_{i=1}^{n} \mathbf{1}_{\{x_i = x\}} \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{x \in \mathcal{X}} \rho(x) \mathbf{1}_{\{x_i = x\}} \\
&= \frac{1}{n} \sum_{i=1}^{n} \rho(x_i).
\end{aligned}$$

Hence,

$$B_n(\rho, D) = \bigcup_{x^n : \frac{1}{n}\sum_{i=1}^n \rho(x_i) \le D} \{x^n\}$$
$$= \bigcup_{p_{x^n} \in \mathbb{P}_n : \langle p, \rho \rangle \le D} T(p)$$

(b) Recall that in class we showed that

$$T(p) \doteq 2^{nH(p)}.$$

So the claim is saying that the size of the union is dominated by the size of the largest set. To prove this, we use the relation in (a) and upper bound

$$\left| \bigcup_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} T(p) \right| \le \sum_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} |T(p)|$$
$$\le |\mathbb{P}^n| \max_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} |T(p)|$$
$$\dot{\le} (n+1)^r \max_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} 2^{nH(p)}$$
$$\dot{\le} 2^{n \max_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} H(p)}.$$

Similarly, for the lower bound we can obtain

$$\left| \bigcup_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} T(p) \right| \ge \max_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} |T(p)|$$
$$\dot{\ge} 2^{n \max_{p \in \mathbb{P}_n : \langle p, \rho \rangle \le D} H(p)}.$$

Along with part (a), this proves the required claim.

(c) Taking $\mathcal{X} = \{0, 1\}$ in (b), and $x^n = y^n$, $\rho(x) = x$ gives the result after noticing that the max in part (b) is given by

$$\max_{q \in [0, D]} h_2(q) = h_2(D)$$

when $D \le 1/2$.

2. In what follows, all random variables have finite alphabet, and all pmfs are defined on finite alphabets.

   (a) Let Z be a random variable with pmf p_z and alphabet $\mathcal{Z}$. Let $T_\delta(p_z)$ be the set of strongly δ-typical sequences with respect to p_z.
   Show that for any $g : \mathcal{Z} \to \mathbb{R}^+$, and $z^n \in T_\delta(p_z)$, we have

$$\left| \frac{1}{n} \sum_{i=1}^n g(z_i) - \mathbb{E}[g(Z)] \right| \le \delta \mathbb{E}[g(Z)].$$

In what follows, it may be useful to invoke part (a) in your arguments.

(b) Let (X, Y) be random variables with joint $p_{x,y}$. Let $T_\delta(p_{x,y})$ be the set of jointly δ-typical sequences with respect to $p_{x,y}$. Show that

$$\frac{1}{n}\sum_{i=1}^{n} d(x_i, y_i) \le (1+\delta)\mathbb{E}[d(X, Y)]$$

for any distortion function $d(x, y)$ and $(x^n, y^n) \in T_\delta(p_{x,y})$.

(c) Let $A_\epsilon(p)$ denote the (weakly) typical set with respect to p, and $T_\delta(p)$ be the set of δ-typical sequences with respect to p. Show that

$$T_\delta(p) \subseteq A_\epsilon(p)$$

for $\epsilon = \delta H(p)$.

(d) Let Q be a distribution with pmf q, and $T_\delta(p)$ be the set of δ-typical sequences with respect to some p. Show that for any $\delta > 0$,

$$Q(T_\delta(p)) \doteq 2^{-n(D(p\|q)-\alpha(\delta))},$$

where $\lim_{\delta\to 0}\alpha(\delta) = 0$.

Solution

(a) From the definition of $T_\delta(p_z)$, we conclude

$$(1-\delta)p_z(z) \le p_{z^n}(z) \le (1+\delta)p_z(z)$$

for all $z \in \mathcal{Z}$. Since

$$\frac{1}{n}\sum_{i=1}^{n} g(z_i) = \sum_{z\in\mathcal{Z}} g(z)p_{z^n}(z),$$

multiplying by $g(z)$ in the first display and summing over $z \in \mathcal{Z}$ gives the result.

(b) Let $Z = (X, Y)$ and apply part (a) with $g(z) = d(x, y)$.

(c) Let $X \sim p$. Applying part (a) to $Z = X$ and $g(z) = -\log(p(z))$ gives

$$(1-\delta)H(p) \le -\frac{1}{n}\sum_{i} p(x_i)\log(p(x_i)) \le (1+\delta)H(p)$$

for any $x^n \in T^\delta(p)$. Rearranging raising to the exponent then gives

$$2^{-n(1+\delta)H(p)} \le p(x^n) \le 2^{-n(1-\delta)H(p)}$$

which implies $x^n \in A_\epsilon(p)$ with $\epsilon = \delta H(p)$.

(d) For the upper bound, we have via a union bound

$$\begin{aligned}
Q(T_\delta(p)) &= Q\left(\bigcup_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} \{x^n : p_{x^n} = \widehat{p}\}\right) \\
&\le \sum_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} Q(T(\widehat{p})) \\
&\dot{\le} \sum_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} 2^{-nD(\widehat{p}\|q)} \\
&\le |\widehat{\mathbb{P}}^n| \max_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} 2^{-nD(\widehat{p}\|q)} \\
&\dot{\le} 2^{-n \min_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} D(\widehat{p}\|q)}.
\end{aligned}$$

For the lower bound, we bound the probability of the union by the probability of the most probable set:

$$\begin{aligned}
Q(T_\delta(p)) &= Q\left(\bigcup_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} \{x^n : p_{x^n} = \widehat{p}\}\right) \\
&\ge \max_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} Q(T(\widehat{p})) \\
&\dot{\ge} \max_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} 2^{-nD(\widehat{p}\|q)} \\
&= 2^{-n \min_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} D(\widehat{p}\|q)}
\end{aligned}$$

which shows that

$$Q(T_\delta(p)) \doteq 2^{-n \min_{\{\widehat{p}\in\mathbb{P}^n : |\widehat{p}-p|\le\delta p\}} D(\widehat{p}\|q)}.$$

Now let p^* be the minimizer in the above display. Then since the type of p^* is in the typical set, we have by (a) that

$$\begin{aligned}
|D(p^*\|q) - D(p\|q)| &= |\mathbb{E}_{p^*}[\log(p^*/q)] - \mathbb{E}_p[\log(p/q)]| \\
&\le |\mathbb{E}_{p^*}[\log(p^*/q)] - \mathbb{E}_{p^*}[\log(p/q)]| + |\mathbb{E}_{p^*}[\log(p/q)] - \mathbb{E}_p[\log(p/q)]| \\
&\le \log(1+\delta) + \delta D(p\|q) \\
&\to 0
\end{aligned}$$

as $\delta \to 0$. Meanwhile, by definition

$$D(p^*\|q) \le D(p\|q),$$

hence we must have

$$D(p^*\|q) = D(p\|q) - \alpha(\delta)$$

where $\alpha \ge 0$ and $\alpha(\delta) \to 0$ as $\delta \to 0$.

3. **Modulo 8 Channel and a Binary Source with Erasure Distortion**
Let a, M be arbitrary real numbers, and let $M > 0$. Then we say, $b = a \bmod \text{M}$, if $b \in [0, M)$ and $b = a - kM$, for some integer k. For example, if $a = 10.2$ and $M = 8$ then $b = 2.2$.
Consider the memoryless channel described by

$$Y = (X + Z) \bmod 8,$$

where the channel input X, output Y, and the noise Z are real valued. The noise is uniformly distributed on $[-B, B]$, i.e. $Z \sim \text{Uniform}[-B, B]$ and is independent of the input X. Assume that $0 < B \leq 4$ is a known channel parameter.

(a) Show that the channel capacity as a function of B is $2 - \log_2 B$.

(b) A communication system is suggested where the permitted channel input values are restricted to the set $\{1, 3, 5, 7\}$. Show that, when $B = 1$, it is still possible to achieve the channel capacity (in fact, to achieve this capacity with zero probability of error and with a very simple scheme).

Consider now a Bernoulli(1/2) source under erasure distortion, i.e., the reconstruction alphabet is $\{0, 1, e\}$ and

$$d(u, v) = \begin{cases} 0 & \text{if } u = v \\ 1 & \text{if } v = e \\ \infty & \text{otherwise} \end{cases}$$

(c) Show that the rate-distortion function of this source is given by $R(D) = 1 - D$ for $0 \leq D \leq 1$.

(d) Can you suggest an explicit scheme (not random coding based) which achieves the optimal rate-distortion performance for the Bernoulli(1/2) source, and distortion D.

(e) Consider now a joint-source-channel-coding setting for communicating the bit source of part (c) through the modulo 8 channel defined in part (a). For every two source symbols we are allotted one channel use, i.e., the encoder translates a block of $2n$ source bits into n channel inputs.
Find the minimal achievable distortion (when $n \to \infty$) as a function of B.

(f) The following communication system is suggested for the setting of the previous part: every two source bits are mapped into a channel input in the following way:

$$\begin{array}{ccc} 00 & \to & 1 \\ 01 & \to & 3 \\ 11 & \to & 5 \\ 10 & \to & 7 \end{array}$$

This mapping is known as a Gray Code. Show that this system (with the corresponding optimal decoder) achieves the minimal distortion in the case where $B = 2$.

Solution: **Modulo 8 Channel and a Binary Source with Erasure Distortion**

(a)
$$I(X;Y) = h(Y) - h(Y|X)$$

Y given X is a shifted (mod 8) version of Z mod 8 and, in turn, Z mod 8 is uniformly distributed on $[0, B] \cup [8 - B, 8]$ so

$$h(Y|X) = h(Z \bmod 8) = h(Z) = \log(2B).$$

Also, Y is supported on $[0, 8]$ so a uniform distribution of Y would maximize $h(Y)$. Is there an input distribution on X that will induce $Y \sim \text{Uniform}[0, 8]$? Of course there is. For example $X \sim \text{Uniform}[0, 8]$ will do the job. To summarize, we got that the maximal value of $I(X;Y)$ across this channel, which is its capacity, is given by

$$\log 8 - \log(2B) = 2 - \log B.$$

(b) For $B = 1$ the capacity we computed in the previous part is 2 bits per channel use. In this case we can communicate 2 bits, error free, with every channel use. We map the 2 bits into the corresponding $x \in \{1, 3, 5, 7\}$. The channel output y is then decoded to the x in that quaternary set closest to it (if there are two such values we can pick one arbitrarily). Since $B = 1$, an error can occur only if $y \in \{0, 2, 4, 8\}$, which is a set of measure 0, hence the probability of error is 0.

(c) For U, V in the feasible set where $Ed(U, V) \le D$ we have

$$\begin{aligned}
I(U;V) &= H(U) - H(U|V) \\
&= 1 - [H(U|V=0)P(V=0) + H(U|V=1)P(V=1) + H(U|V=e)P(V=e)] \\
&= 1 - H(U|V=e)P(V=e) \\
&\ge 1 - P(V=e) \\
&\ge 1 - D,
\end{aligned}$$

where the last equality is due to the fact that $H(U|V=0) = H(U|V=1) = 0$ under any pair in the feasible set for this distortion criterion, the first inequality to the fact that U is binary, and the second inequality to the fact that $P(V = e) = Ed(U, V) \le D$. Can we find a joint distribution in the feasible set that will satisfy the two inequalities with equality? The following construction does the job: let $B \sim \text{Bernoulli}(D)$ be independent of U, and let V be given by

$$V = \begin{cases} e & \text{if } B = 1 \\ U & \text{if } B = 0 \end{cases}$$

(d) For block length n, let the reconstruction codebook consist of all sequences in $\{0, 1, e\}^n$ for which the last nD entries are e and the first $n(1 - D)$ entries are 0 or 1. Clearly the size of the codebook is $2^{n(1-D)}$ and the rate is $1 - D$. The encoder maps each sequence in $\{0, 1\}^n$ to the closest element in the codebook, e.g., for $n = 4$, $D = 0.5$, the codebook consists of $00ee, 01ee, 10ee, 11ee$ and the reconstruction of 0010 is $00ee$. The reconstructions is e in the last nD places and hence the average distortion is D. Thus, this scheme achieves rate $1 - D$ when $0 \leq D \leq 1$ and is optimum.

(e) According to the JSCC theorem, the maximal number of source symbols per channel use we can communicate at distortion level D is

$$\rho(D) = \frac{2 - \log B}{1 - D}, \qquad 0 \leq D \leq 1.$$

The minimal achievable distortion in question is the D for which $\rho(D) = 2$, which is readily solved here to be given by $\frac{1}{2} \log B$, provided $B \geq 1$. For $B < 1$ we already have $\rho(0) > 2$ so can easily achieve distortion 0 at rate 2. in summary, the minimum achievable distortion is $\max\left\{\frac{1}{2} \log B, 0\right\}$.

(f) For $B = 2$ the minimum distortion computed in the previous part is 1/2. Consider pairs of source symbols that are adjacent to each other in the Gray Code (modulo 8, so that 1 and 7 are also adjacent). Note that such pairs are always in agreement on one bit, and are not in agreement on the other bit. Therefore, when $B = 2$, we can always reconstruct one bit with certainty, and for the other bit there is not one value more likely than the other, and so we reconstruct it as an erasure. In other words, we employ the following reconstruction scheme:

$$\begin{array}{lcl} y \in [0,1] \cup [7,8] & \rightarrow & e0 \\ y \in [1,3] & \rightarrow & 0e \\ y \in [3,5] & \rightarrow & e1 \\ y \in [5,7] & \rightarrow & 1e \end{array}$$

It is readily verified that with this scheme one gets average distortion [across the two source symbols encoded] of 1/2, with probability 1. According to the preceding part, the minimum distortion attainable is 1/2, so this simple scheme indeed attains optimum performance.