ECE 788: Network Information Theory
Midterm 2009

Please provide *clear and detailed* answers.
The points associated to each question are indicative (6 points are expected to be enough to obtain a sufficient score, 10 to obtain the maximum score).
The set $\{1, 2, ..., N\}$ is denoted below as $[1, N]$

1. (3 points) Pick a sequence $X^n \in \mathcal{X}^n$ according to a uniform distribution in the set of all sequences $\mathcal{X}^n$ with $\mathcal{X} = \{0, 1\}$.
1.1. What is the probability of any sequence $X^n = x^n$, i.e., $\Pr[X^n = x^n]$?

Sol.:

$$\Pr[X^n = x^n] = 2^{-n},$$

since we have 2^n sequences overall.

1.2. Find and upper and a lower bound on the probability that X^n falls in the set $T_0^n(P_X)$ for some type P_X, i.e., $\Pr[X^n \in T_0^n(P_X)]$.

Sol.:

$$\begin{aligned} \Pr[X^n &\in T_0^n(P_X)] = \sum_{x^n \in T_0^n(P_X)} \Pr[X^n = x^n] \\ &= |T_0^n(P_X)| 2^{-n}, \end{aligned}$$

so that

$$\frac{2^{n(H(P_X)-1)}}{(n+1)^2} \leq \Pr[X^n \in T_0^n(P_X)] \leq 2^{n(H(P_X)-1)}.$$

1.3. Find an upper bound on the probability that X^n falls in the set $\mathcal{S}$ of *all* sequences $x^n \in \mathcal{X}^n$ whose type P_X has entropy less or equal than R bits, i.e., $\Pr[X^n \in \mathcal{S}]$.

Sol.:

$$\begin{aligned} \Pr[X^n &\in \mathcal{S}] = \sum_{P_X\colon H(P_X) \leq R} \Pr[X^n \in T_0^n(P_X)] = \sum_{P_X\colon H(P_X) \leq R} 2^{n(H(P_X)-1)} \\ &\leq (\text{number of types with } H(P_X) \leq R) \cdot 2^{n(R-1)} \end{aligned}$$

where we have used the fact that $2^{n(H(P_X)-1)} \leq 2^{n(R-1)}$. Now, the number of types is known to be always less or equal to $(n+1)^2$, so that we have

$$\Pr[X^n \in \mathcal{S}] \leq (n+1)^2 \cdot 2^{n(R-1)}.$$

1.4. Using the result above, calculate an upper bound to

$$\lim_{n \to \infty} \frac{\log_2 \Pr[X^n \in \mathcal{S}]}{n}$$

and conclude from this that, for n large enough, $\Pr[X^n \in \mathcal{S}] \lesssim 2^{n(R-1)}$. This says that for $R < 1$, we have $\Pr[X^n \in \mathcal{S}] \to 0$ as $n \to \infty$. Can you interpret this result by drawing a connection with the AEP?

Sol.: We have

$$\begin{aligned} & \lim_{n\to\infty} \frac{\log \Pr[X^n \in \mathcal{S}]}{n} \\ \leq & \lim_{n\to\infty} \frac{2\log(n+1)}{n} + (R-1) \\ = & R-1. \end{aligned}$$

Drawing sequences uniformly from $X^n \in \mathcal{X}^n$ is statistically equivalent to drawing X^n with i.i.d. letters distributed according to the uniform distribution U in $\mathcal{X} = \{0,1\}$ ($U(0) = U(1) = 1/2$). Therefore, we can use the AEP to conclude that X^n lies in the typical set $T^n_\epsilon(U_X)$ with high probability, but $T^n_\epsilon(U_X)$ is included in S only if $R = 1$.

2. (3 points) We want to prove achievability of the lossless coding problem in a way alternative to the one seen in class. Consider the usual i.i.d. source with pmf P_X. We are interested in analyzing the performance of the following compression scheme.
Code construction: For each source sequence $x^n \in \mathcal{X}^n$ randomly and independently draw an integer number taken uniformly from the set $[1, 2^{nR}]$ for some $R > 0$. Such index is called the *bin index* of x^n and denoted as $f(x^n)$. Notice that $f(x^n) \in [1, 2^{nR}]$ and that, in general, many sequences x^n are assigned the same bin index.
Encoding: Given a source sequence x^n, the encoder outputs the bin index $w = f(x^n)$. Notice that w consists of R bits/ symbol.
Decoding: The decoder reconstructs a sequence $\hat{x}^n \in \mathcal{X}^n$ such that $\hat{x}^n \in T_\epsilon(P_X)$ *and* $f(\hat{x}^n) = w$. If it can find none or more than one such $\hat{x}^n$, it declares an error.
We want to prove that if $R > H(X)$ we can find a coding scheme in the class described above (i.e., a binning function f) such that the probability of error $P_e^n(f) = \Pr[\hat{X}^n \neq X^n | f]$ tends to zero as $n \to \infty$. To do this, we use random coding arguments and evaluate the average error probability $P_e^n = E_f[P_e^n(f)]$ with respect to the bin function f.
2.1. Identify the two error events $\mathcal{E}_0$ and $\mathcal{E}_1$, and, using the law of total expectation, write the average probability of error P_e^n in terms of the error events (Hint: The decoder makes a mistake when $x^n \notin T_\epsilon(P_X)$ and when we can find a $\hat{x}^n \neq x^n$ which satisfies the required conditions at the decoder).

Sol.: We have

$$\begin{aligned} \mathcal{E}_0 &= \{X^n \notin T_\epsilon(P_X)\} \\ \mathcal{E}_1 &= \bigcup_{\substack{\hat{x}^n \neq X^n, \\ \hat{x}^n \in T^n_\epsilon(P_X)}} \{f(\hat{x}^n) = f(X^n)\} \end{aligned}$$

and the probability if error is given by: $P_e^n = \Pr[\mathcal{E}_0 \cup \mathcal{E}_1] = \Pr[\mathcal{E}_0] + \Pr[\mathcal{E}_1 | \mathcal{E}_0]$.

2.2. Using the result above, specify the steps (and theorems) necessary to upper bound the average probability of error P_e^n as

$$P_e^n \leq \delta_\epsilon(n) + 2^{nH(X)(1+\epsilon)} \cdot \Pr[f(\hat{x}^n) = w],$$

where $\Pr[f(\hat{x}^n) = w]$ is the probability that the bin index assigned to sequence $\hat{x}^n$ is equal to any specific index w.

Sol.: We have $\Pr[\mathcal{E}_0] \leq \delta_\epsilon(n)$ by the AEP. Moreover,

$$\begin{aligned}
\Pr[\mathcal{E}_1|\mathcal{E}_0] &= \frac{\sum_{x^n \in T_\epsilon^n(P_X)} p(x^n) \Pr[\mathcal{E}_1|X^n = x^n]}{\Pr[X^n \in T_\epsilon^n(P_X)]} \\
&= \Pr[\mathcal{E}_1|X^n = x^n \in T_\epsilon^n(P_X)] \frac{\sum_{x^n \in T_\epsilon^n(P_X)} p(x^n)}{\Pr[X^n \in T_\epsilon^n(P_X)]} \\
&= \Pr[\mathcal{E}_1|X^n = x^n \in T_\epsilon^n(P_X)] \\
&\leq \sum_{\substack{\hat{x}^n \neq x^n, \\ \hat{x}^n \in T_\epsilon^n(P_X)}} \Pr[f(\hat{x}^n) = f(x^n)] \\
&= \Pr[f(\hat{x}^n) = w] \cdot |T_\epsilon^n(P_X)| \\
&\leq \Pr[f(\hat{x}^n) = w] \cdot 2^{nH(X)(1+\epsilon)},
\end{aligned}$$

where in the second and fifth lines we have used the symmetry of the bin function generation, in the fourth the union bound and in the last the AEP.

2.3. Justify the relationship $\Pr[f(\hat{x}^n) = w] = 2^{-nR}$. Based on this, conclude the proof (i.e., show that there exists a coding scheme...).

3. (2 points) We want to prove that the function $C(S)$ (capacity vs. cost) is concave in S, that is, that $C(\lambda S_1 + (1-\lambda)S_2) \geq \lambda C(S_1) + (1-\lambda)C(S_2)$ for any $0 \leq \lambda \leq 1$ and $S_1, S_2 \geq 0$.
3.1. Illustrate this relationship with a sketch.
3.2. Argue that proving this relationship is equivalent to showing that there exist at least a scheme with cost $E[s^n(X^n)] \leq \lambda S_1 + (1-\lambda)S_2$ that achieves a rate larger or equal than $\lambda C(S_1) + (1-\lambda)C(S_2)$ (Hint: Remember the meaning of capacity!)

Sol.: This is because the capacity is the rate of the best possible scheme, so that it cannot be worse of that of any specific strategy.

3.3. Using the idea of time-sharing between two schemes, one achieving $C(S_1)$ and the other achieving $C(S_2)$, find one such scheme. You have to show that the proposed scheme achieves the required rate and that it satisfies the cost constraint.

Sol.: Use the scheme that achieves $C(S_1)$ for λn channel uses and the other scheme for the remaining $(1-\lambda)n$ channel uses. The rate is given by

$$R = \frac{\lambda n C(S_1) + (1-\lambda) n C(S_2)}{n} = \lambda n C(S_1) + (1-\lambda) n C(S_2).$$

since the first scheme sends $C(S_1)$ bits per channel use and similarly $C(S_2)$ the cost is

$$
\begin{aligned}
E[s^n(X^n)] &= \frac{1}{n}\sum_{i=1}^{n} E[s(X_i)] \\
&= \frac{1}{n}\left(\sum_{i=1}^{\lambda n} E[s(X_i)] + \sum_{i=(1-\lambda)n+1}^{n} E[s(X_i)]\right) \\
&= \lambda\frac{1}{\lambda n}\sum_{i=1}^{\lambda n} E[s(X_i)] + (1-\lambda)\frac{1}{(1-\lambda)n}\sum_{i=(1-\lambda)n+1}^{n} E[s(X_i)] \\
&= \lambda S_1 + (1-\lambda)S_2,
\end{aligned}
$$

by definition of cost for the two schemes.

4. (2 points) Consider a discrete *memoryless* channel $P_{Y|X}$ over alphabets $\mathcal{X}$ and $\mathcal{Y}$. We generate a codebook $\mathcal{C}$ by drawing each codeword $X^n(w)$, with $w \in [1, 2^{nR}]$ for some rate $R > 0$, i.i.d. and independently with pmf P_X. Without loss of generality, consider transmission of message $w = 1$. In the following, we prove the channel coding theorem in an alternative way.

4.1. Given a received signal $Y^n = y^n$, define the list $\mathcal{L}(y^n)$ as

$$\mathcal{L}(y^n) = \{w \in [2, 2^{nR}]\colon (x^n(w), y^n) \in T_\epsilon(P_{XY}))\}.$$

Argue that the average probability of error for a receiver based on joint typicality, given a received signal y^n, is upper bounded by

$$\delta_\epsilon(n) + \Pr[|\mathcal{L}(y^n)| \geq 1],$$

using the law of total expectation. Then, use an appropriate inequality (which one?) to show that

$$\Pr[|\mathcal{L}(y^n)| \geq 1] \leq E[|\mathcal{L}(y^n)|].$$

Sol.: Define $\mathcal{E}_0 = \{(X^n(1), Y^n) \notin T^n_\epsilon(P_{XY})\}$

$$
\begin{aligned}
P^n_e &= \Pr[\mathcal{E}_0] + \Pr[\mathcal{L}(y^n) \geq 1|\mathcal{E}_0^c]\Pr[\mathcal{E}_0^c] \\
&\leq \delta_\epsilon(n) + \Pr[\mathcal{L}(y^n) \geq 1],
\end{aligned}
$$

where in the second line we have used the AEP and the fact that the codewords of different messages are generated independently.

4.2. Finally, prove that $E[|\mathcal{L}(y^n)|] \to 0$ as $n \to \infty$ if $R < I(X;Y)$, which concludes the proof of the channel coding theorem (why?).

Sol.:

$$
\begin{aligned}
E[|\mathcal{L}(y^n)|] &= E\left[\sum_{\hat{w}\neq 1} 1((X^n(\hat{w}), y^n) \in T_\epsilon(P_{XY}))\right] \\
&= \sum_{\hat{w}\neq 1} \Pr[(X^n(\hat{w}), y^n) \in T_\epsilon(P_{XY})] \\
&\leq 2^{nR} \cdot 2^{-n(I(X;Y)-2\epsilon H(X))},
\end{aligned}
$$

which follows from the fundamental lemma and concludes the proof.

5. (1.5 points) Consider transmission over a channel with two inputs X_1 and X_2 and two outputs Y_1 and Y_2, where the channel is given by $P_{Y_1Y_2|X_1X_2} = P_{Y_1|X_1} \cdot P_{Y_2|X_2}$ with $P_{Y_i|X_i}(y|x)$ being Binary Symmetric Channel (BSC) with probability of bit flipping p_i, $i = 1, 2$. No cost constraint is imposed. Find the capacity and the optimal input distribution $P_{X_1X_2}$ (Hint: Recall that the joint entropy is always less or equal than the sum of the marginal entropies).

Sol.:

$$\begin{aligned} C &= \max_{P_{X_1X_2}} I(X_1X_2; Y_1Y_2) \\ &\leq \max_{P_{X_1X_2}} H(Y_1) + H(Y_2) - H(Y_1Y_2|X_1X_2) \\ &= \max_{P_{X_1}} H(Y_1) + \max_{P_{X_1}} H(Y_2) - H(p_1) - H(p_2) \\ &\leq 2 - H(p_1) - H(p_2), \end{aligned}$$

which is achieved with $P_{X_1X_2} = P_{X_1}P_{X_2}$ and $P_{X_1}(0) = P_{X_1}(1) = 1/2$, $i = 1, 2$.

6. (1.5 points) Consider two sequences X^n and Y^n generated i.i.d. so that

$$X_i = Y_i \oplus Z_i, \; i = 1, 2, ..., n$$

with Y^n i.i.d. with $\Pr[Y_i = 1] = 1/2$ and Z^n i.i.d. with $\Pr[Z_i = 1] = p$, independent. Assume that the encoder sees both X^n and Y^n and, based on both sequences, produces an index $w \in [1, 2^{nR}]$ for some rate R. The decoder receives w and measures also Y^n. Based on w and Y^n, the decoder wants to recover X^n losslessly, i.e., $\Pr[\hat{X}^n \neq X^n] \to 0$ for $n \to \infty$. Find an achievable rate R. (Hint: Propose a scheme based on the fact that Y^n is known at both encoder and decoder and on joint typicality arguments).

Sol.: Since Y^n is known at both sides, the encoder just needs to send Z^n! The rate is then $R = H(p)$.
More generally, the encoder can assign a different index w to all sequences in the set $T_\epsilon^n(P_{XY}|y^n)$ and a different unique w to all other sequences. Notice that, since both encoder and decoder know y^n, such set is known at both encoder and decoder. Due to the conditional AEP the probability of error goes to zero as $n \to \infty$. Moreover, the number of indices needed is

$$2^{nR} = \lceil |T_\epsilon^n(P_{XY}|y^n)| \rceil + 1 \leq 2^{nH(X|Y)(1+\epsilon)} + 2,$$

from which it follows that an achievable rate is given for n large by:

$$R = H(X|Y) = H(p).$$