ECE 776 - Information theory (Spring 2011)
Final

1. (2 points) Consider the source coding problem, and assume that, unlike the usual scenario, the decoder is not informed a priori about the decoding function $\hat{X}^l = g(W)$, where l is the block size and W is the encoder-to-decoder message of rate R bits/source sample.
a. Argue that in order for the encoder to inform the decoder about $g(W)$, the encoder needs to send

$$2^{lR} \log_2 \left(|\hat{\mathcal{X}}|^l \right) \text{ bits,}$$

where $\hat{\mathcal{X}}$ is the reconstruction alphabet, in addition to the message W (Hint: How many functions $g(W)$ are there?).
b. Consider now compressing an i.i.d. source X^n. Assume that the encoder knows the source pmf $p(x)$, but the decoder does not (which implies that it does not even know the decoding function). Argue that, by partitioning the source in blocks of size $l = \log_2 \log_2 n$, lossless compression is still possible at the same rate $R = H(X)$ that one would need if $p(x)$ was known also to the decoder (Hint: The encoder needs to inform the decoder about the decoding function. Moreover, we have that, if $l = \log_2 \log_2 n$, then $\lim_{n\to\infty} \frac{2^{lR}\log_2(|\hat{\mathcal{X}}|^l)}{n} = 0$).

2. (2 points) The histogram $q_{x^n}(a)$ of a sequence $x^n \in \mathcal{X}^n$ is the function

$$q_{x^n}(a) = \frac{|\{i\colon\ x_i = a\}|}{n},$$

where $a \in \mathcal{X}$.
a. Show that if X^n is i.i.d. with pmf $p(x)$, we have

$$p(x^n) = 2^{-n(H(q_{x^n}) + D(q_{x^n}||p))}.$$

(Hint: $p(x^n) = \Pi_{i=1}^n p(x_i)$. Also, multiplying and dividing by a given term may be useful).
b. Using the point above, show that the number of sequences with a given histogram $q_{x^n}(a)$ is at most $2^{nH(q_{x^n})}$ (Hint: Recall the proof of the AEP).
c. It is known that the number of all possible histograms is no larger than $(n+1)^{|\mathcal{X}|}$. Using this fact, argue that, even if neither the encoder nor the decoder know $p(x)$, compression needs only rate $H(q_{x^n})$ for a sequence x^n with histogram q_{x^n}.

3. (2 points) Consider a binary symmetric channel (BSC) with bit flipping probability p, n channel uses and cost constraint $\frac{1}{n}\sum_{i=1}^n E[X_i] \leq q$.
a. Find the capacity.
b. Given a binary source V^m, i.i.d. Ber(s), find the minimal bandwidth ratio $b = \frac{n}{m}$ necessary for transmission of V^m over the given BSC with a Hamming distortion of D (You can use known results from rate-distortion theory).
c. Consider $b = 1$, $q = s = 1/2$, and a simple joint source-channel coding scheme with $X_i = V_i$. Find the resulting Hamming distortion and argue that this scheme is optimal.

4. (2 points) Consider the pdf $f(x) = \exp(-x)$, $x \geq 0$.

a. Find the typical set $A_\epsilon^{(n)}$ (You can use natural log in the definition of typical set).
b. Draw the set for $n = 2$.
c. Using the AEP, what can we say about the volume of $A_\epsilon^{(n)}$?
d. Consider quantizing this source symbol by symbol with a uniform quantizer with step Δ and then using a standard entropy encoder on the resulting discrete sequence. If Δ is sufficiently small, what is the rate required by this scheme?

5. (2 points) Consider an i.i.d. source uniformly distributed in the set $\{1, ..., m\}$. Find the rate-distortion function for the Hamming distortion (Hint: Use the Fano inequality and then look for a $p(x, \hat{x})$ that satisfies the bound so obtained. Symmetry helps).

6. (2 points) Let $p(y|x)$ be a discrete memoryless channel with capacity C. Suppose that the output y goes through an erasure channel $p(s|y)$ that erases each symbol with probability α, so that $s = y$ with probability $1 - \alpha$ and $s = e$ (e is the erasure symbol) with probability α. Find the capacity. (Hint: Define a new random variable Z, where $Z = 1$ if $S = e$ and $Z = 0$ otherwise. Note that knowing Z does not bring any information about Y).

Sol.:

1.
a. In order to communicate function $g(W)$, the encoder needs to describe a sequence $\hat{\mathcal{X}}^l$ for each one of the 2^{lR} messages W. Therefore, the total number of functions $g(W)$ is

$$(\hat{\mathcal{X}}^l)^{2^{lR}},$$

and thus the total number of bits is $\log_2\left((\hat{\mathcal{X}}^l)^{2^{lR}}\right) = 2^{lR}\log_2(\hat{\mathcal{X}}^l)$.
b. We partition the n source samples in blocks of size $l = \log_2\log_2 n$, and encode each one independently. We thus need $lH(X)$ bits per block and $nH(X)$ bits overall to send the messages, one for each block. For the decoder to be able to decode, we also need to communicate the decoding function. Overall, we thus need to send rate

$$\frac{1}{n}\left(nH(X) + 2^{lR}\log_2(\hat{\mathcal{X}}^l)\right),$$

but, since $\lim_{n\to\infty}\frac{2^{lR}\log_2\left(|\hat{\mathcal{X}}|^l\right)}{n} = 0$, we have that the rate is $H(X)$.

2.
a. We have

$$p(x^n) = \Pi_{i=1}^n p(x_i) = \Pi_{a\in\mathcal{X}} p(a)^{nq_{x^n}(a)},$$

so that

$$\begin{aligned}
\frac{1}{n}\log_2 p(x^n) &= \sum_{a\in\mathcal{X}} q_{x^n}(a)\log_2 p(a) \\
&= \sum_{a\in\mathcal{X}} q_{x^n}(a)\log_2\left(p(a)\frac{q_{x^n}(a)}{q_{x^n}(a)}\right) \\
&= \sum_{a\in\mathcal{X}} q_{x^n}(a)\log_2\left(\frac{p(a)}{q_{x^n}(a)}\right) - H(q_{x^n}) \\
&= -D(q_{x^n}(a)||p(a)) - H(q_{x^n}),
\end{aligned}$$

from which the result follows.
b. Consider an i.i.d. sequence with pmf given by $q_{x^n}(a)$. Denoting as $\mathcal{A}$ the set of sequences with histogram $q_{x^n}(a)$. We have that the probability that the given sequence is in $\mathcal{A}$ is

$$|\mathcal{A}|2^{-nH(q_{x^n})} \leq 1,$$

from which the result follows.
c. It is known that the number of possible histograms is no larger than $(n+1)^{|\mathcal{X}|}$. Therefore, the encoder can describe the histogram to decoder with $|\mathcal{X}|\log_2(n+1)$ bits, and then send the index of the specific sequence with the given histogram with $nH(q_{x^n})$ bits. Therefore, even if neither the encoder nor the decoder know $p(x)$, compression needs rate

$$\frac{1}{n}\left(|\mathcal{X}|\log_2(n+1) + nH(q_{x^n})\right) \to H(q_{x^n}) \text{ for } n \to \infty.$$

3. a. The capacity is given by

$$C = \max_{\pi \leq q} H(Y) - H(p) = \begin{cases} H(p*q) - H(p) & \text{if } q \leq 1/2 \\ 1 - H(p) & \text{if } q \geq 1/2 \end{cases}$$

where $p*q = p(1-q) + (1-p)q$ and $\Pr[X=1] = \pi$.
b. We have

$$b = \frac{R(D)}{C} = \frac{H(s) - H(D)}{C} \text{ if } D \leq \min\{s, 1-s\},$$

whereas $b = 0$ if $D \geq \min\{s, 1-s\}$.
c. The received signal is

$$Y_i = V_i + Z_i,$$

where Z_i is Ber(p). Therefore, the Hamming distortion is p.
Now from the equation above, we have that the minimum distortion is given by

$$1 = \frac{1 - H(D)}{1 - H(p)},$$

so that $D = p$ is the minimum distortion.

4. (2 points) a. For the pdf $f(x) = \exp(-x)$, $x \geq 0$, we have $h(X) = \log e = 1$ nats, so

$$\begin{aligned} A_\epsilon^{(n)} &= \left\{ x^n \in \mathbb{R}^n \colon x_i \geq 0 \text{ and } \left| \frac{1}{n}\sum_{i=1}^n x_i - 1 \right| \leq \epsilon \right\} \\ &= \left\{ x^n \in \mathbb{R}^n \colon x_i \geq 0 \text{ and } 1 - \epsilon \leq \frac{1}{n}\sum_{i=1}^n x_i \leq 1 + \epsilon \right\}. \end{aligned}$$

b. The set for $n = 2$ is simply $A_\epsilon^{(2)} = \left\{x^2 \in \mathbb{R}^2 \colon x_i \geq 0 \text{ and } 1 - \epsilon \leq \frac{1}{2}(x_1 + x_2) \leq 1 + \epsilon\right\}$, which is easily drawn.
c. From the AEP, the volume of $A_\epsilon^{(n)}$ satisfies $vol(A_\epsilon^{(n)}) \leq 2^n$ for all n, and $vol(A_\epsilon^{(n)}) \geq (1-\epsilon)2^n$ for sufficiently large n.

d. The rate required by the given scheme is $h(X) - \log_2 \Delta = \log_2 e - \log_2 \Delta$ bits/source sample.

5. We have

$$R(D) = \min_{\substack{p(\hat{x}|x): \\ \Pr[X \neq \hat{X}] \leq D}} I(X; \hat{X}).$$

First, note that $R(D) = 0$ if $D \geq 1 - 1/m$, which is obtained by setting $\hat{X}$ equal to any number in $\{1, ..., m\}$. Then, assume $D < 1 - 1/m$ and set $\Pr[X \neq \hat{X}] = a \leq D$. We get

$$I(X; \hat{X}) = \log_2 m - H(X|\hat{X}),$$

where using Fano inequality

$$H(X|\hat{X}) \leq H(\Pr[X \neq \hat{X}]) + \Pr[X \neq \hat{X}] \log_2(m-1),$$

so that

$$\begin{aligned} I(X; \hat{X}) &\geq \log_2 m - H(a) - a \log_2(m-1) \\ &\geq \log_2 m - H(D) - D \log_2(m-1), \end{aligned}$$

The question is now if this can be achieved by some $p(x, \hat{x})$. Consider the natural choice where

$$p(x, \hat{x}) = \begin{cases} \frac{1-D}{m} & \text{for } x = \hat{x} \\ \frac{D}{m(m-1)} & \text{otherwise} \end{cases}.$$

One can easily check that this satisfies the condition $p(x) = 1/m$ and is such that

$$\begin{aligned} I(X; \hat{X}) &= \log_2 m - H(X|\hat{X}) \\ &= \log_2 m - H([1 - D, \frac{D}{m-1}, ..., \frac{D}{m-1}]) \\ &= \log_2 m - H(D) - D \log_2(m-1). \end{aligned}$$

6. The capacity of the channel is

$$\max_{p(x)} I(X; S).$$

Note that Z is a function of S and is independent of X. Therefore, we have

$$\begin{aligned} I(X; S) &= I(X; SZ) \\ &= I(X; Z) + I(X; S|Z) \\ &= 0 + \Pr[Z = 1] I(X; S|Z = 1) + \Pr[Z = 0] I(X; S|Z = 0) \\ &= (1 - \alpha) I(X; S|Z = 1) + \alpha \cdot 0 \\ &= (1 - \alpha) I(X; S|Z = 1) \\ &= (1 - \alpha) I(X; Y), \end{aligned}$$

so that the capacity is

$$\max_{p(x)} I(X; S) = (1 - \alpha) C$$

where C is the capacity of the channel $p(y|x)$.