ECE 1502 — Information Theory
November 16, 2007

7.9 *Suboptimal codes.* From the proof of the channel coding theorem, it follows that using a random code with codewords generated according to probability $p(x)$, we can send information at a rate $I(X;Y)$ corresponding to that $p(x)$ with an arbitrarily low probability of error. For the Z channel described in the previous problem, we can calculate $I(X;Y)$ for a uniform distribution on the input. The distribution on Y is (3/4, 1/4), and therefore

$$I(X;Y) = H(Y) - H(Y|X) = H(\frac{3}{4}, \frac{1}{4}) - \frac{1}{2}H(\frac{1}{2}, \frac{1}{2}) = \frac{3}{2} - \frac{3}{4}\log 3. \quad (1)$$

7.14 *Channels with dependence between the letters.*

(a) First,

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= H(X_1, X_2) - H(X_1, X_2|Y_1, Y_2) \\ &= H(X_1, X_2), \end{aligned}$$

where the last equality follows from the fact that (Y_1, Y_2) uniquely identifies (X_1, X_2).

(b)

$$\begin{aligned} C &= \max_{p(x_1,x_2)} H(X_1, X_2) \\ &= \log_2(4) = 2, \end{aligned}$$

when the four input pairs are used equiprobably.

(c) When the input pairs are equiprobable, it is easy to show that

$$p(x_1|y_1) = p(x_1).$$

Therefore,

$$\begin{aligned} I(X_1; Y_1) &= H(X_1) - H(X_1|Y_1) \\ &= H(X_1) - H(X_1) \\ &= 0. \end{aligned}$$

7.19 *Capacity of the carrier pigeon channel.*

(a) Since a pigeon arrives at the destination every 5 minutes, there are 12 pigeons per hour that arrive safely. Each pigeon carries an 8 bit message, thus the capacity of the link is 96 bits/hour.

Effectively, we get to use an errorless channel (with 256 inputs and 256 outputs, and thus capacity $C_1 = \log_2(256) = 8$ bits/channel-use) 12 times per hour.

(b) We can model this problem using an erasure channel with 256 inputs and 257 outputs (256 output symbols that match the inputs, and one erasure); the input is received correctly with probability $1-\alpha$, and an erasure occurs with probability α.

Just an in the binary-input erasure channel, it is easy to show that the capacity is

$$C_2 = (1-\alpha)\log_2(256) = 8(1-\alpha) \text{ bits/channel-use},$$

which is equivalent to $96(1-\alpha)$ bits/hour.

(c) We can model this problem using a 256-ary symmetric channel (just like the BSC, but with 256 inputs and 256 outputs). A pigeon arrives safely with probability $1-\alpha$, and if the pigeon is shot down, the probability that the dummy carries the intended message is $\frac{1}{256}$. Therefore, the message is received correctly with probability

$$1-\alpha+\alpha(\frac{1}{256}) = 1-\frac{255\alpha}{256}.$$

Furthermore, when an error occurs (i.e., the pigeon gets shot, and the dummy carries the wrong message), all of the incorrect messages are equiprobable.

Finally, by symmetry, it is clear that the uniform input distribution is capacity-achieving. Let X represent the channel input, and Y represent the channel output. Then, for equiprobable inputs,

$$\begin{aligned} C_3 &= H(Y) - H(Y|X) \\ &= \log_2(256) - H(1-\frac{255\alpha}{256}, \frac{\alpha}{256}, \frac{\alpha}{256}, \ldots, \frac{\alpha}{256}) \\ &= 8 - H_2(\frac{255\alpha}{256}) - \frac{255\alpha}{256}\log_2(255), \end{aligned}$$

and the link capacity is $12C_3$ bits/hour.

7.27 *Erasure channel.*

Expanding $I(X;Y,S)$ in two ways,

$$\begin{aligned} I(X;Y,S) &= I(X;Y) + I(X;S|Y) \\ &= I(X;S) + I(X;Y|S). \end{aligned}$$

Now, since $X \to Y \to S$, we have $I(X;S|Y) = 0$, and thus

$$\begin{aligned} I(X;S) &= I(X;Y) - I(X;Y|S) \\ &= I(X;Y) - \alpha H(Y|S=e) + \alpha H(Y|X, S=e) \\ &= H(Y) - H(Y|X) - \alpha \log_2|\mathcal{Y}| + \alpha H(Y|X). \end{aligned}$$

Therefore,

$$C = \max_{p(x)} H(Y) - (1-\alpha)H(Y|X) - \alpha\log_2|\mathcal{Y}|,$$

where the maximizing input distribution is, in general, a function of $p(y|x)$ and α.

7.35 *Capacity.*

(a) Due to the structure of $\hat{\mathcal{P}}$ (in particular, the off-diagonal zeros), we can interpret $\hat{\mathcal{P}}$ to be made up of two distinct channels with non-overlapping inputs and outputs, and at each channel use, the user must communicate using one of the two channels. In other words, the user selects a channel (and thus conveys info by doing so), and then transmits a symbol over the given channel (conveying more info).

Therefore, following our solution to Problem 7.28 from Assignment #3, we know that the effective capacity is

$$C_{eff} = \log_2(2^C + 2^{C_a}),$$

where C is the capacity of $\mathcal{P}$, and C_a is the capacity of the noiseless unary channel. Clearly, $C_a = 0$, therefore

$$C_{eff} = \log_2(2^C + 1).$$

(b) Similarly, since I_k represents a noiseless k-ary channel, its capacity is $C_a = \log_2 k$, and we have

$$C_{eff} = \log_2(2^C + k).$$

8.1 *Differential Entropy.*

(a) Exponential distribution.

$$\begin{aligned} h(f) &= -\int_0^\infty \lambda e^{-\lambda x}[\ln\lambda - \lambda x]dx && (2)\\ &= -\ln\lambda + 1 \text{ nats.} && (3)\\ &= \log\frac{e}{\lambda} \text{ bits.} && (4) \end{aligned}$$

(b) Laplace density.

$$\begin{aligned} h(f) &= -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|}[\ln\frac{1}{2} + \ln\lambda - \lambda|x|]\,dx && (5)\\ &= -\ln\frac{1}{2} - \ln\lambda + 1 && (6)\\ &= \ln\frac{2e}{\lambda} \text{ nats.} && (7)\\ &= \log\frac{2e}{\lambda} \text{ bits.} && (8) \end{aligned}$$

(c) Sum of two normal distributions.

The sum of two normal random variables is also normal, $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, and we have

$$h(f) = \frac{1}{2}\log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.} \qquad (9)$$

8.7 *Differential entropy bound on discrete entropy.*

Note: The solution to this problem is from Section 9.7 of Cover & Thomas (First Edition).

As in the hint, we have $Y = X' + U$, and therefore the distribution of Y has the shape of a histogram (that is, $f_Y(y) = p_i$ for $i \le y < i + 1$). It is clear that $H(X') = H(X)$, since discrete entropy depends only on the probabilities and not on the values of the outcomes. Now

$$\begin{aligned}
H(X') &= -\sum_{i=i}^{\infty} p_i \log_2 p_i \\
&= -\sum_{i=i}^{\infty} \left(\int_i^{i+1} f_Y(y)dy \right) \log_2 \left(\int_i^{i+1} f_Y(y)dy \right) \\
&= -\sum_{i=i}^{\infty} \int_i^{i+1} f_Y(y) \log_2 f_Y(y) dy \\
&= -\int_1^{\infty} f_Y(y) \log_2 f_Y(y) dy \\
&= h(Y),
\end{aligned}$$

since $f_Y(y) = p_i$ for $i \leq y < i+1$.

Hence we have the following chain of inequalities:

$$\begin{aligned}
H(X) &= H(X') \\
&= h(Y) \\
&\leq \frac{1}{2} \log_2 (2\pi e)(\mathrm{Var(Y)}) \\
&= \frac{1}{2} \log_2 (2\pi e)(\mathrm{Var(X')} + \mathrm{Var(U)}) \\
&= \frac{1}{2} \log_2 (2\pi e) \left(\sum_{i=1}^{\infty} p_i i^2 - \left(\sum_{i=1}^{\infty} p_i i \right)^2 + \frac{1}{12} \right).
\end{aligned}$$

Since entropy is invariant with respect to permutation of $p_1, p_2, \ldots$, we can also obtain a bound by a permutation of the p_i's. We conjecture that a good bound on the variance will be achieved when the high probabilities are close together, i.e., by the assignment $\ldots, p_5, p_3, p_1, p_2, p_4, \ldots$ for $p_1 \geq p_2 \geq \ldots$.

8.8 *Channels with uniformly distributed noise.*

$$\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} h(Y) - h(Y|X) \\
&= \max_{p(x)} h(Y) - h(X+Z|X) \\
&= \max_{p(x)} h(Y) - h(Z) \\
&= \max_{p(x)} h(Y) - \log_2 2,
\end{aligned}$$

where in the last line we have used the fact that the differential entropy of a random variable that is distributed uniformly between α and $\alpha + a$ is $\log_2 a$ bits.

Furthermore, we see that the output of the channel, Y, is limited to values in the range $[-3, 3]$. From a result on distributions with maximum entropy (specifically, see Chapter 12, Example 12.2.4), we see that $h(Y)$ will be maximized if we select $p(x)$ such that the distribution of Y is uniform in the range $[-3, 3]$.

Now, for

$$p(x = 0) = p(x = 2) = p(x = -2) = \frac{1}{3},$$

it is easy to see that the distribution of Y is uniform in the range $[-3, 3]$, and from our previous discussion, $h(Y) = \log_2 6$.

Therefore, we have

$$C = \log_2 6 - \log_2 2 = \log_2 3.$$

Note that one could have arrived at this result using calculus (without explicit knowledge that the uniform distribution maximizes entropy for a variable with bounded range), but it would have involved lengthy (and tedious) calculations.

8.10 *Shape of the typical set.*

Since the X_i are i.i.d., $f(x^n) = c^n e^{-(x_1^4+x_2^4+\cdots+x_n^4)}$. From the definition of the typical set, x^n is typical if and only if

$$2^{-n(h+\epsilon)} \leq c^n e^{-(x_1^4+x_2^4+\cdots+x_n^4)} \leq 2^{-n(h-\epsilon)},$$

which is equivalent to

$$-n(h+\epsilon)\ln 2 \leq n \ln c - (x_1^4 + x_2^4 + \cdots + x_n^4) \leq -n(h-\epsilon)\ln 2.$$

Finally, re-arranging the preceding condition, we have

$$A_\epsilon^{(n)} = \{x^n \in \mathcal{R}^n | n(\ln C + (h-\epsilon)\ln 2) \leq x_1^4 + x_2^4 + \cdots + x_n^4 \leq n(\ln C + (h+\epsilon)\ln 2).\}$$

This is reminiscient of the fact that the typical set of a Gaussian distribution is a thin spherical shell; in this case, the sphere has been replaced by a shape of the form $x_1^4 + x_2^4 + \cdots + x_n^4 = r$, but the 'thin shell' property persists.

9.4 *Exponential noise channels.*

As specified in Problem Set 4, we assume $X_i \geq 0$ for all i, since otherwise the capacity is unbounded when the only constraint is on the mean of the input (i.e., we could have unbounded power). Now, since

$$EX_i \leq \lambda,$$

it follows that

$$EY_i \leq \mu + \lambda.$$

Furthermore, from the non-negativity constraints on X_i and Z_i, we also have $Y_i \geq 0$ for all i.

Since the noise is independent of the input, we have

$$\begin{aligned} I(X_i;Y_i) &= h(Y_i) - h(Y_i|X_i) \\ &= h(Y_i) - h(X_i + Z_i|X_i) \\ &= h(Y_i) - h(Z_i). \end{aligned}$$

Now, since Y_i is a non-negative random variable with $EY_i \leq \mu + \lambda$, it follows from Example 12.2.4 (Cover & Thomas) that $h(Y_i)$ is maximized when Y_i is exponentially distributed. Furthermore, from Problem 8.1, we know that the differential entropy of an exponential distribution is a monotonically increasing function of the mean. Therefore,

$$h(Y_i) \leq \log_2(e(\mu + \lambda)),$$

and it follows that

$$C \leq \log_2(e(\mu + \lambda)) - \log_2(e\mu)) = \log_2(1 + \frac{\lambda}{\mu}).$$

To confirm that $C = \log_2(1 + \frac{\lambda}{\mu})$, we need to show that X_i can be chosen such that $X_i + Z_i$ has an exponential distribution with mean $\mu + \lambda$. We will do so by explicitly deriving the required distribution.

The characteristic function of an exponential random varible Z with mean ρ is

$$\Phi_Z(\omega) = \frac{\frac{1}{\rho}}{\frac{1}{\rho} - j\omega}.$$

Since

$$\Phi_Y(\omega) = \Phi_X(\omega)\Phi_Z(\omega),$$

we have that the required $\Phi_X(\omega)$ is given by

$$\begin{aligned} \Phi_X(\omega) &= \frac{1 - j\mu\omega}{1 - j(\mu + \lambda)\omega} \\ &= \frac{1}{1 - j(\mu + \lambda)\omega} - \left(\frac{\mu}{\mu + \lambda}\right)\left(\frac{1}{1 - j(\mu + \lambda)\omega}\right) + \frac{\mu}{\mu + \lambda}. \end{aligned}$$

Finally, taking the inverse transform, we have

$$f_X(x) = \left(\frac{\lambda}{\mu + \lambda}\right)\left(\frac{e^{-\frac{x}{\mu+\lambda}}}{\mu + \lambda}\right) + \frac{\mu}{\mu + \lambda}\delta(x), \text{ for } x \geq 0$$

Observe that this solution specifies that we should transmit zero with probability $\mu/(\mu + \lambda)$, and otherwise transmit an exponentially distributed nonzero value with as large a mean value as possible. (In a sense we are "saving up" for the possibility of a "louder" transmission whenever we send zero.)