Noisy Channel Coding

Given a channel defined by conditional probabilities $p(y|x)$ of output y given input x, how can we use it to communicate reliably? Shannon's idea was to consider the following scheme:

$$x \longrightarrow \boxed{\text{channel}} \longrightarrow y$$

$$m \longrightarrow \boxed{\text{encoder}} \xrightarrow{X} \boxed{\text{channel}} \xrightarrow{Y} \boxed{\text{decoder}} \longrightarrow m'$$

Use an encoder and decoder to transmit messages m reliably.

In the iid case (memoryless channels), we can easily prove that messages can be reliably transmitted at rate $r \leq \max_{P_X} I(X:Y)$ by using block encoders and decoders. That is, for blocks of length n (with $n \to \infty$), roughly $\ell = nr$ bits can be sent with negligible probability of error.

The encoding and decoding strategy revolves around typical sets. Define the encoder to be a function f from M to X^n; later we will consider, in usual Shannon fashion, random choices of encoding function. To be concrete, let the random variable M be uniformly distributed on ℓ bits. The decoder receives the output y of the channel and determines the set of inputs m' such that $f(m')$ is in the conditionally-typical set $T_{n,\varepsilon}(X|Y=y)$ for the observed y.

What can go wrong in this scheme? Two things: 1. y may not be typical, in which case the conditionally-typical set is empty, or 2. there's more than one input m for which $f(m)$ is conditionally typical. Call these errors E_1 and E_2. By the union bound, the total error probability $P_e(m,f)$ for a given message m and code f is bounded by

$$P_e(m,f) \leq P(E_1|m,f) + P(E_2|m,f)$$

Let's now average over all possible encodings f, so that we can view the encoder as choosing a random x, distributed according to P_X, for any input m. Importantly, we still have the freedom to choose P_X. Now we have

$$P_e(m) \leq P(E_1|m) + P(E_2|m)$$

The first term is less than δ because y is unlikely to be nontypical when drawing from the distribution $P_Y(Y=y) = \sum_x P_{XY}(X=x, Y=y)$

The second term is somewhat more involved. E_2 occurs if $\exists\, m' \neq m$ such that $f(m') \in T_{n,\varepsilon}(X|Y=y)$. The probability for this is

$$P[E_2|m] \leq \sum_{m' \neq m} P\left(f(m') \in T_{n,\varepsilon}(X|Y=y)\right) = \sum_{m' \neq m} \sum_{x \in T_{n,\varepsilon}(X|Y=y)} P_X(x)$$

$$\leq 2^{\ell}\, 2^{-n(H(X)-\varepsilon)}\, |T_{n,\varepsilon}(X|Y=y)| \leq 2^{\ell - n(H(X) - H(X|Y) - 3\varepsilon)}$$

$$= 2^{\ell - n(I(X:Y) - 3\varepsilon)}$$

Thus, if we choose $\ell = n(I(X:Y) - 4\varepsilon)$, the probability of error given message m will be small.

This shows that the error probability is small, averaging over all possible encodings of M into random variable X^n. What we'd really like is to know that there exists an encoding such that the maximum probability of error (maximizing over messages m) is small.

We can show this in the following roundabout two-step procedure. First, average over all messages m. Then there must exist at least one code for which the average (over m) error probability is small, say 2δ (In fact, most of them have this property, by the Markov ineq.)

Next, we can throw out the worst half of the codewords m: rank the m in terms of error probability and then throw out the worst half. The resulting $2^{\ell-1}$ messages all have an error probability less than 4δ

$$\max_{m\in \underline{M}} P_e(m) \leq \min_{m\in\overline{M}} P_e(m) \leq 2^{-\ell+1}\sum_{m\in\overline{M}} P_e(m) \leq 2\,\frac{1}{2^\ell}\sum_m P_e(m) \leq 2\cdot 2\delta = 4\delta$$

We have to go through this procedure and cannot directly remove the average over encoding functions at the level of $P[E_2|m]$ because that doesn't tell us the maximum over m for a given code.

We have shown that $r = I(X:Y) - 4\frac{\varepsilon}{n}$ is achieveable. Since we're free to choose P_X, this is immediately extended to $r = \max_{P_X} I(X:Y) - 4\frac{\varepsilon}{n}$

In the limit $n\to\infty$, block coding therefore achieves the capacity

$$C = \max_{P_X} I(X:Y)$$

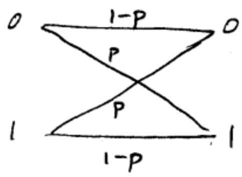
Converse

- Upper bound on capacity. Any encoding/decoding scheme gives rise to a Markov chain $M \to X^n \to Y^n \to M'$. Successful decoding means $M=M'$ and therefore $H(M)=I(M:M')$. But from the information processing inequality $I(A:C) \leq I(A:B)$ for Markov chain $A \to B \to C$. Applying this to $M \to Y^n \to M'$ we have $I(M:M') \leq I(M:Y^n)$. Then Markov chain $M \leftarrow X^n \leftarrow Y$ gives $I(M:Y^n) \leq I(X^n:Y^n)$, and thus $H(M) \leq I(X^n:Y^n) = nI(X:Y)$. Since M is uniformly distributed, $H(M)=\ell$ and therefore $\ell \leq nI(X:Y)$. Maximizing over P_X gives $r < C$.

- error probability for $r > C$.

Example: Binary symmetric channel:

input bit is flipped with prob. p



$p(0|0) = 1-p \qquad p(1|0) = p$

$p(0|1) = p \qquad p(1|1) = 1-p$

Let $\Pr(X=0) = q$. Then $p(x,y) = \begin{pmatrix} q(1-p) & pq \\ p(1-q) & (1-p)(1-q) \end{pmatrix}$

$H(X) = h_2(q) \qquad h_2(x) = -x\log x - (1-x)\log(1-x)$

$H(XY) = -q(1-p)\log q(1-p) - pq\log pq - p(1-q)\log p(1-q) - (1-p)(1-q)\log(1-p)(1-q)$

$= -(1-p)\log(1-p) - q\log q - p\log p - (1-q)\log(1-q)$

$= h_2(p) + h_2(q)$

$H(Y) = -[q(1-p)+p(1-q)]\log[q(1-p)+p(1-q)] - [pq+(1-p)(1-q)]\log[pq+(1-p)(1-q)]$

$I(X:Y) = H(X) + H(Y) - H(XY)$

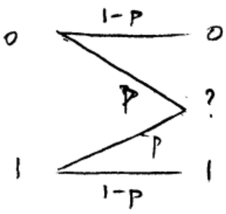
$= H(Y) - h_2(p)$

$\partial_q I(X:Y) = \partial_q H(Y)$ but entropy is concave, so optimal $q = \frac{1}{2}$

$H(Y) = 1$, $H(XY) = 1 + h_2(p)$, $H(X) = 1$

$C = I(X:Y) = 1 - h_2(p)$

Erasure channel



capacity: $1-p$

try $q = \frac{1}{2}$ $\quad H(X) = 1 \quad H(Y) = -(1-p)\log\frac{1-p}{2} - p\log p = h_2(p) + (1-p)$

$H(XY) = -(1-p)\log\frac{1-p}{2} - p\log\frac{p}{2} = h_2(p) + 1 \qquad I(X:Y) = 1-p$

consider feedback from the receiver, which can only help: receiver tells sender which bits are erased. Sender resends. Scheme has rate $1-p$, so forward capacity $C < 1-p$.