Ph/CS 219
2 March 2011
Jointly typical sequences

Here is a little more detail about the proof of the Noisy Channel Coding Theorem, specifically the proof that the mutual information is an achievable rate. (Based on Chapter 8 of Cover and Thomas.)

A noisy classical channel is characterized by the conditional probability function p(y|x), the probability that the letter y is received when the letter x is sent. We consider using the channel n times to send n letters. We use a code with rate R; that is we send one of 2^{nR} n-letter messages, so that we attempt to convey nR bits of information in n uses of the channel. We say that the rate R is achievable if there is a sequence of codes with rate R such that the probability of of a decoding error approaches zero as n -> infinity. The capacity of the channel is the supremum of achievable rates.

Following Shannon, we consider constructing an n-letter code by generating 2^{nR} codewords, each time sampling from an i.i.d. probability distribution, in which the letter x is selected with probability p(x). For any such code, we consider a codeword selected uniformly at random from among the 2^{nR} possible codewords. We would like to obtain an upper bound on the probability of a decoding error when this codeword is sent through n uses of the channel. To do that we have to choose and analyze a decoding procedure.

Our decoding procedure is "jointly typical decoding". When a correlated probability distribution p(x,y) is sampled n times to generate strings

$$(x_1 x_2 \ldots x_n,\ y_1 y_2 \ldots y_n) \equiv (\bar{x}, \bar{y}),$$

by the law of large numbers we know that for each fixed $\delta, \varepsilon > 0$ we can choose n sufficiently large so that, with probability $> 1-\varepsilon$ $(\bar{x},\bar{y})$ is "jointly δ-typical", i.e.

$$2^{-n[H(X)+\delta]} \le p(\bar{x}) \le 2^{-n[H(X)-\delta]},$$

$$2^{-n[H(Y)+\delta]} \le p(\bar{y}) \le 2^{-n[H(Y)-\delta]},$$

$$2^{-n[H(XY)+\delta]} \le p(\bar{x},\bar{y}) \le 2^{-n[H(XY)-\delta]}.$$

The number N_{typ} of jointly typical sequences satisfies

$$1 \geq \sum_{typ} p(\bar{x},\bar{y}) \geq N_{typ}\, 2^{-n[H(XY)+\delta]}$$

$$\Longrightarrow \quad N_{typ} \le 2^{n[H(XY)+\delta]}$$

When Bob receives $\bar{y}$, then if there is unique codeword $\bar{x}$ jointly typical with $\bar{y}$ he decodes $\bar{y}$ as $\bar{x}$. Otherwise he decodes in an arbitrary way.

A decoding error occurs if either of the following happen:

(1) The sent and received messages are not jointly typical. This occurs with probability $\leq \varepsilon$.

(2) There is a codeword $\bar{x}$ other than the one sent which is jointly typical with the received message $\bar{y}$.

Suppose that $\bar{x}_1$ was actually sent and $\bar{y}_1$ received, and let $\bar{x}_2$ be another codeword different than $\bar{x}_1$. What is the probability that $\bar{x}_2$ and $\bar{y}_1$ are jointly typical?

Because $\bar{x}_2$ and $\bar{y}_1$ were determined by sampling independently, they are uncorrelated: the probability that $\bar{x}_2$ and $\bar{y}_1$ were generated factorizes into $p(\bar{x}_2)\, p(\bar{y}_1)$ (the product of the marginal distributions for $\bar{x}$ and $\bar{y}$). The probability of joint typicality is

$$\sum_{(\bar{x}_2,\bar{y}_1)\ \mathrm{typ}} p(\bar{x}_2)\, p(\bar{y}_1) \leq \underbrace{2^{n[H(XY)+\delta]}}_{\text{upper bound on } N_{typ}}\ \underbrace{2^{-n[H(X)-\delta]}}_{\text{upper bound on prob. of typical } \bar{x}_2}\ \underbrace{2^{-n[H(Y)-\delta]}}_{\text{upper bound on prob. of typical } \bar{y}_1}$$

$$\leq 2^{-n[I(X;Y)-3\delta]}$$

There are $(2^{nR}-1)$ codewords other than $\bar{x}_1$ that might have been sent. So, averaged over codes and codewords,

$$\text{Prob. of decoding error} \leq \varepsilon + (2^{nR}-1)\, 2^{-n[I(X;Y)-3\delta]}$$

$$\leq \varepsilon + 2^{-n[I-R-3\delta]} \longrightarrow 0 \quad \text{as } n\to\infty$$

for any $R < I$

Slepian-Wolf coding

In Sec. 5.1.2 of the lecture notes, it is claimed that if a joint distribution p(x,y) is sampled n times, where Alice receives n-letter message x and Bob receives n-letter message y, then Alice can send nH(X|Y) bits to Bob, enabling Bob to determine x with high asymptotic success probability. Here we explain in more detail the coding scheme that Alice and Bob use to achieve this. It is a special case of "Slepian-Wolf coding" (Cover and Thomas Sec. 14.4).

Alice sorts all possible n-letter messages into 2^{nR} bins which are selected uniformly at random. The choice of bins is known to both Alice and Bob. Alice sends to Bob the nR bits that identify the bin that contains her message x. Thus Bob knows both y and the bin; he decodes y as x if x is the unique message in this bin that is jointly typical with y. Otherswise he choses an arbitrary decoding.

A decoding error occurs if
(1) The Alice's message x and Bob's message y are not jointly typical. This occurs with probability no larger than epsilon.

If $(\bar{x},\bar{y})$ are jointly typical, then

$$p(\bar{x}|\bar{y}) = \frac{p(\bar{x},\bar{y})}{p(\bar{y})} \geq \frac{2^{-n[H(XY)+\delta]}}{2^{-n[H(Y)-\delta]}} = 2^{-n[H(X|Y)+2\delta]}$$

If y is typical, let $S(X|\bar{y})$ denote the set of $\bar{x}$ that are jointly typical with $\bar{y}$. Then

$$1 \geq \sum_{\bar{x}\in S(X|y)} p(\bar{x}|\bar{y}) \geq |S(X|y)|\, 2^{-n[H(X|Y)+2\delta]}$$

The no. of elements of $S(X|y)$

$$\Longrightarrow |S(X|y)| \leq 2^{n[H(X|Y)+2\delta]}$$

Because the bins are chosen uniformly at random, each $\bar{x}$ is contained in a particular specified bin with probability 2^{-nR}. The probability that $\bar{x}$ is in the bin containing Alice's message by accident is

$$\leq 2^{-nR}|S(X|y)|$$

$$\leq 2^{-n[R-H(X|Y)-2\delta]} \longrightarrow 0 \text{ as } n\to\infty$$

for $R > H(X|Y)$.

Coding in which Alice sends $H(X|Y)$ bits per letter is achievable. (If ave. over codes has decoding error probability $<\varepsilon$, then there is a particular code with error probability $<\varepsilon$.)