Northwestern University
Department of Electrical and Computer Engineering

Lecture 14: Supplemental notes

The following are notes on random coding bound for DMC's. These are based primarily on material from Chapter 5 of [2].

1 Introduction

In the previous lecture we presented a proof of the direct part of the coding theorem for a discrete memoryless channel (DMC). This proof follows that in Section 8.7 of [1] and was based on the idea of using a random coding argument with *typical set* decoding. Typical set decoding allowed us to use the properties of jointly typical sequences to bound the probability of decoding error for large enough block lengths. Using this approach we proved that for *large enough* block lengths, n, the probability of error can be made arbitrarily small; however, this proof gives no indication of how large a block length is required before the error probability can be made acceptably small. In other words, we have no indication of how large n must be for these asymptotic results to be meaningful, or equivalently, how fast the error probability goes to zero with increasing block length. For DMC's (and many other cases) it can be shown that the error probability goes to zero exponentially fast (in the block length) at rates below capacity. Results of this type are sometimes referred to as *strong coding theorems*; a result, as proved last time, that simply states that rates less than capacity are achievable is called a *weak coding theorem*.

In these notes, we will discuss the following result for a DMC with (information) capacity C.

Random coding bound: *For all rates $R < C$, there exists a sequence of $(2^{nR}, n)$ block codes with average probability of error $P_e^{(n)}$ that satisfies*

$$P_e^{(n)} \leq 2^{-nE_r(R)},$$

where $E_r(R) > 0$ for all $R < C$.

The quantity $E_r(R)$ is called the *random coding exponent* for the channel. It gives an upper-bound on the exponential rate at which the probability of error goes to zero. Note that this bound is true for all n.

For a rate $R < C$, define $P_e(2^{nR}, n)$ to be the minimum probability that can be attained by any $(2^{nR}, n)$ block code for the given channel. The *reliability function* of the channel is defined to be

$$E(R) = \lim_{n\to\infty} \frac{-\log P_e(2^{nR}, n)}{n}.$$

In other words, the reliability function indicates the fastest rate at which the probability of error goes to zero. For a DMC, for large enough rates R, it can be shown that $E(R) = E_r(R)$. In some special cases, such as a binary erasure channel, the reliability function is known for all rates $R < C$. For a general DMC, upper and lower bounds on the reliability function are known, but the exact reliability function is not known for all rates, R.

2 Random coding bound

Consider a discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, p(y|x))$. In this section we give a sketch of the main ideas used to prove the random coding bound.

As in the proof of the coding theorem, we again consider an ensemble of randomly chosen codes, where each symbol in each codeword is chosen i.i.d. using a given p.m.f.[1] $p(x)$ on $\mathcal{X}$. For a given $(2^{nR}, n)$ code, let $x^n(m)$ denote the mth codeword, for $m = 1, \ldots, M$, where $M = 2^{nR}$. As before, the p.m.f. for the mth codeword is given by

$$p(x^n(m)) = \prod_{i=1}^{n} p(x_i),$$

where $x^n(m) = (x_1, \ldots, x_n)$.

Instead of jointly typical decoding, we now consider maximum likelihood (M.L.) decoding. That is if y^n is the received sequence, the decoder chooses the message m such that

$$p(y^n|x^n(m)) \geq p(y^n|x^n(m')) \; \forall m' \neq m.$$

(ties can be broken arbitrarily.) Assuming all codewords are equally likely, this is equivalent to MAP decoding, which minimizes the probability of error.

Using M.L. decoding, once again let $\Pr(\mathcal{E}|W = 1)$ denote the average probability of decoding error when the 1st message is transmitted, averaged over the entire ensemble of random codes. As in the last lecture, if all messages are equally likely, then the overall average probability of error, $\Pr(\mathcal{E})$, averaged over all possible random codes is equal to $\Pr(\mathcal{E}|W = 1)$. Also, as in the proof of the coding theorem it follows that at least one code in the ensemble must have a average probability of error less than $\Pr(\mathcal{E}) = \Pr(\mathcal{E}|W = 1)$. We first show the following bound on this quantity.

Lemma 1: *For any* $\rho \in [0, 1]$,

$$\Pr(\mathcal{E}|W = 1) \leq (M-1)^{\rho} \sum_{y^n \in \mathcal{Y}^n} \left[\sum_{x^n \in \mathcal{X}^n} p(x^n) p(y^n|x^n)^{1/(1+\rho)} \right]^{1+\rho}.$$

[1] In the following, we use the imprecise notation $p(x)$ to denote the p.m.f. of a random variable X, where it is understood that the argument x indicates the random variable. Hence, $p(y)$ denotes a different p.m.f. than $p(x)$.

Proof: The average error probabilty can be written as

$$\Pr(\mathcal{E}|W=1) = \sum_{x^n(1)\in\mathcal{X}^n} p(x^n(1)) \sum_{y^n\in\mathcal{Y}^n} p(y^n|x^n(1)) \Pr[error|x^n(1), y^n].$$

Here $p(x^n(1))$ indicates the probability that $x^n(1)$ is choosen as the 1st codeword, and $\Pr[error|x^n(1), y^n]$ indicates the probability of error given $x^n(1)$ is transmitted as the 1st codeword and y^n is received. This event will depend on the choice of the other codewords in the code.

Given that $x^n(1)$ is transmitted and y^n is received, define the events $A_{m'}$ for all $m' \neq 1$, as the event that the codeword $x^n(m')$ is selected to be in the code and has a likelihood that is larger that $x^n(1)$, i.e. $p(y^n|x^n(m')) \geq p(y^n|x^n(1))$. Then,

$$\Pr[error|x^n(1), y^n] \leq \Pr\left(\bigcup_{m'\neq 1} A'_m\right).$$

This term can be upper bounded by the union bound, but this will not give us the desired result. Instead we use a related bound, i.e.,

$$\Pr\left(\bigcup_{m'\neq 1} A'_m\right) \leq \left[\sum_{m'\neq 1} \Pr(A'_m)\right]^\rho \qquad \text{for any } \rho \text{ such that } 0 < \rho \leq 1.$$

To see that this is true, consider the following two cases. First if $S = \sum_{m'\neq 1} \Pr(A'_m) < 1$, then $S^\rho > S$ and the bound follows from the union bound. Otherwise, if $S \geq 1$ then $S^\rho \geq 1$, and the bound follows since the term on the left is a probability.

Next, for any $m' \neq 1$ we have

$$\begin{aligned}\Pr(A'_m) &= \sum_{\{x^n : p(y^n|x^n) \geq p(y^n|x^n(1))\}} p(x^n) \\ &\leq \sum_{x^n\in\mathcal{X}^n} p(x^n) \frac{p(y^n|x^n)^s}{p(y^n|x^n(1))^s} \quad \text{for any } s > 0.\end{aligned}$$

Combining the above, we have

$$\begin{aligned}\Pr[error|x^n(1), y^n] &\leq \left[\sum_{m'=2}^{M} \sum_{x^n} p(x^n) \frac{p(y^n|x^n)^s}{p(y^n|x^n(1))^s}\right]^\rho, \\ &= \left[(M-1)\sum_{x^n} p(x^n) \frac{p(y^n|x^n)^s}{p(y^n|x^n(1))^s}\right]^\rho.\end{aligned}$$

Next, averaging over all possible values for $x^n(1)$ and y^n, we have

$$\begin{aligned}\Pr(\mathcal{E}|W=1) &\leq \sum_{x^n(1)}\sum_{y^n} p(x^n(1))p(y^n|x^n(1))\left[(M-1)\sum_{x^n} p(x^n)\frac{p(y^n|x^n)^s}{p(y^n|x^n(1))^s}\right]^\rho \\ &= (M-1)^\rho \sum_{y^n}\left[\sum_{x^n(1)} p(x^n(1))p(y^n|x^n(1))^{1-s\rho}\right]\left[\sum_{x^n} p(x^n)p(y^n|x^n)^s\right]^\rho.\end{aligned}$$

If we set $s = \frac{1}{1+\rho}$ and recognize that $x^n(1)$ and x^n are just dummy variables for the summation, we have

$$\Pr(\mathcal{E}|W=1) \leq (M-1)^\rho \sum_{y^n}\left[\sum_{x^n} p(x^n)p(y^n|x^n)^{1/(1+\rho)}\right]^{1+\rho}.$$

as desired. ■

So far we have not used that the channel is a discrete memoryless channel; if this is the case then

$$p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i),$$

Likewise, since each letter of the codewords are generated i.i.d., we have

$$p(x^n) = \prod_{i=1}^n p(x_i).$$

Using these in the above bound gives

$$\begin{aligned}\Pr(\mathcal{E}) &\leq (M-1)^\rho \sum_{y_1}\cdots\sum_{y_n}\left[\sum_{x_1}\cdots\sum_{x_n}\prod_{i=1}^n p(x_i)p(y_i|x_i)^{1/(1+\rho)}\right]^{1+\rho} \\ &= (M-1)^\rho \sum_{y_1}\cdots\sum_{y_n}\left[\prod_{i=1}^n\sum_{x} p(x)p(y_i|x)^{1/(1+\rho)}\right]^{1+\rho} \\ &= (M-1)^\rho \prod_{i=1}^n\sum_{y}\left[\sum_{x} p(x)p(y|x)^{1/(1+\rho)}\right]^{1+\rho}.\end{aligned}$$

The last two lines each follow from writting out the sums, factoring each term, and recognizing that the above summations are the same for each x_i and y_i.

Recall that $M = 2^{nR}$ is the number of codewords, thus

$$\begin{aligned}\Pr(\mathcal{E}) &\leq 2^{nR\rho}\prod_{i=1}^n\sum_{y}\left[\sum_{x} p(x)p(y|x)^{1/(1+\rho)}\right]^{1+\rho} \\ &= 2^{-n[E_0(\rho,p(x))-\rho R]},\end{aligned}$$

where

$$E_0(\rho, p(x)) \triangleq -\log\left(\sum_y \left(\sum_x p(x)p(y|x)^{1/(1+\rho)}\right)^{1+\rho}\right).$$

This bound on the probability of error is true for any input distribution $p(x)$ and any $\rho \in [0,1]$ (and any blocklength n). Thus the tightest bound can be found by optimizing over these parameters, this gives the random coding exponent, $E_r(R)$. Specifically, define

$$E_r(R) = \sup_{0\le\rho\le 1} \sup_{p(x)} \left[E_0(\rho, p(x)) - \rho R\right]. \tag{1}$$

It then follows that

$$\Pr(\mathcal{E}) \le 2^{-nE_r(R)},$$

and so there must exsit some specific $(2^{nR}, n)$ code, for each n, with

$$P_e^{(n)} \le 2^{-nE_r(R)},$$

as stated above.

It can be shown that $E_r(R)$ is positive for all rate $R < C$, this along with the above bound provides an alternative proof of the direct part of the channel coding theorem (see [2]). It can also be shown that $E_r(R)$ is decreasing in R, this reflects the intuition that the error probability can be made smaller by lowering the transmission rate. As $R \to C$, $E_r(R) \to 0$.

3 Cut-off rate

A looser bound on the error probability can be found by setting $\rho = 1$, in this case the optimization in (1) becomes

$$\sup_{p(x)} \left[E_0(1, p(x)) - R\right] = \left(\sup_{p(x)} E_0(1, p(x))\right) - R.$$

This leads to the definition of the "cut-off" rate, R_0, as

$$R_0 = \left(\sup_{p(x)} E_0(1, p(x))\right).$$

This gives the following simpler bound on the probability of error,

$$\bar{P}_{e,m} \le 2^{-n(R_0 - R)},$$

which is meaningful for $R < R_0$. It can be shown that R_0 is strictly less than C. Note that setting $\rho = 1$ is equivalent to using the union bound in the proof of Lemma 1.

At one-time R_0 was thought to be an indicator of the rate that can be achieved over a channel using "practical" coding techniques - this was motivated in part by an analysis of *sequential decoding* techniques for convolutional codes. This has since been shown to be incorrect and practical coding schemes for rates higher than R_0 have been found.

References

[1] T. Cover and J. Thomas, "Elements of Information Theory," Wiley, New York, 1981.

[2] R. Gallager, "Information Theory and Reliable Communication," Wiley, New York, 1968.