ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 18 Information Theory and Coding
Notes on Random Coding December 12, 2003

Random Coding

In this note we prove the achievability part of the channel coding theorem for discrete memoryless channels without feedback. The discussion is largely based on chapter 5 of R. G. Gallager, *Information Theory and Reliable Communication,* Wiley, 1968.

1. Discrete Memoryless Channels Without Feedback

Throughout this note we will fix the channel we wish to communicate over. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets of this channel. We will assume that the channel is discrete, i.e., that $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. The behavior of the channel will be completely described by specifying for each $k \geq 1$ the function

$$P_k : \mathcal{X}^k \times \mathcal{Y}^k \to \mathbb{R}, \quad (x_1^k, y_1^k) \mapsto P_k(y_k|x_1^k, y_1^{k-1}),$$

which gives the probability of receiving the letter y_k at the output at time k, given the past and current inputs to the channel, and the past outputs of the channel.

We make the two following assumptions: the channel is *memoryless* and used *without feedback*.

We will call a channel memoryless if

$$P_k(y_k|x_1^k, y_1^{k-1}) = P(y_k|x_k),$$

which in words means that the channel keeps no memory of the *past* inputs and outputs in determining the output of time k. Also note that the channel does not behave differently at different times: the function P on the right hand side is not an explicit function of k.

If a memoryless channel is used *without feedback*, i.e., if

$$P_{X_k|X_1^{k-1},Y_1^{k-1}}(x_k|x_1^{k-1}, y_1^{k-1}) = P_{X_k|X_1^{k-1}}(x_k|x_1^{k-1})$$

(in words: if the channel inputs do not depend on the past channel outputs) then

$$\begin{aligned}
P_{Y_1^n|X_1^n}(y_1^n|x_1^n) &= \frac{P_{X_1^n,Y_1^n}(x_1^n, y_1^n)}{P_{X_1^n}(x_1^n)} \\
&= \frac{\prod_{k=1}^n P_{X_k,Y_k|X_1^{k-1},Y_1^{k-1}}(x_k, y_k|x_1^{k-1}, y_1^{k-1})}{P_{X_1^n}(x_1^n)} \\
&= \frac{\prod_{k=1}^n P_{X_k|X_1^{k-1},Y_1^{k-1}}(x_k|x_1^{k-1}, y_1^{k-1}) P_{Y_k|X_1^k,Y_1^{k-1}}(y_k|x_1^k, y_1^{k-1})}{P_{X_1^n}(x_1^n)} \\
&= \frac{\prod_{k=1}^n P_{X_k|X_1^{k-1}}(x_k|x_1^{k-1}) P_{Y_k|X_k}(y_k|x_k)}{P_{X_1^n}(x_1^n)} \\
&= \prod_{k=1}^n P(y_k|x_k)
\end{aligned}$$

where we use the memoryless and without feedback conditions at the fourth equality.

From now on, we will restrict our attention of channels used without feedback. With some abuse of notation we will let $P(\mathbf{y}|\mathbf{x})$ denote the probability of receiving the sequence $\mathbf{y} = y_1^n$ at the output of the channel when the channel input is the sequence $\mathbf{x} = x_1^n$. If the channel is memoryless, we see from above that

$$P(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{n} P(y_k|x_k).$$

2. Block Codes

A block code with M messages and block length n is a mapping from a set of M messages $\{1, \ldots, M\}$ to channel input sequences of length n. Thus, a block code is specified when we specify the M channel input sequences $\mathbf{c}_1 = (c_{1,1}, \ldots, c_{1,n}), \ldots, \mathbf{c}_M = (c_{M,1}, \ldots, c_{M,n})$ the messages are mapped into. We will call $\mathbf{c}_m$ the codeword for message m.

To send message m with such a block code we simply give the sequence $\mathbf{c}_m$ to the channel as input.

A decoder for such a block code is a mapping from channel output sequences $\mathcal{Y}^n$ to the set of M messages $\{1, \ldots, M\}$. For a given decoder, let $D_m \subset \mathcal{Y}^n$ denote the set of channel outputs which are mapped to message m. Since an output sequence $\mathbf{y}$ is mapped to exactly one message, D_m's form a collection of disjoint sets whose union is $\mathcal{Y}^n$.

We define the rate of a block code with M messages and block length n as

$$\frac{\ln M}{n},$$

and given such a code and a decoder we define

$$P_{e,m} = \sum_{\mathbf{y} \notin D_m} P(\mathbf{y}|\mathbf{c}_m),$$

the probability of a decoding error when message m is sent. Further define

$$P_{e,\text{ave}} = \frac{1}{M} \sum_{m=1}^{M} P_{e,m} \quad \text{and} \quad P_{e,\max} = \max_{1 \le m \le M} P_{e,m}$$

as the average and maximal (both over the possible messages) error probability of such a code and decoder.

Among many possible decoding methods, the rule that minimizes $P_{e,\text{ave}}$ is the maximum likelihood rule. Given a channel output sequence $\mathbf{y}$, the maximum likelihood rule decodes a message m for which

$$P(\mathbf{y}|\mathbf{c}_m) \ge P(\mathbf{y}|\mathbf{c}_{m'}) \quad \text{for every } m' \ne m,$$

and if there are more than one such m chooses one of them arbitrarily. We will restrict ourselves in the following to the maximum likelihood rule.

3. Error probability for two codewords

Consider now the case when $M = 2$, so the block code consists of two codewords, $\mathbf{c}_1$ and $\mathbf{c}_2$. We will find a bound on $P_{e,m}$ for the maximum likelihood decoding rule.

Suppose message 1 is to be sent. The channel input is then $\mathbf{c}_1$, and the probability of receiving $\mathbf{y}$ at the channel output is $P(\mathbf{y}|\mathbf{c}_1)$. An error will occur if the received sequence is not in D_1. Since for every $\mathbf{y}$ not in D_1, $P(\mathbf{y}|\mathbf{c}_2) \geq P(\mathbf{y}|\mathbf{c}_1)$, (but there may be $\mathbf{y}$'s for which $P(\mathbf{y}|\mathbf{c}_2) = P(\mathbf{y}|\mathbf{c}_1)$ and $\mathbf{y} \in D_1$) we have

$$\begin{aligned}
P_{e,1} &= \sum_{\mathbf{y}\notin D_1} P(\mathbf{y}|\mathbf{c}_1) \\
&\leq \sum_{\mathbf{y}:\, P(\mathbf{y}|\mathbf{c}_2)\geq P(\mathbf{y}|\mathbf{c}_1)} P(\mathbf{y}|\mathbf{c}_1) \\
&= \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{c}_1)\mathbb{1}_{P(\mathbf{y}|\mathbf{c}_2)\geq P(\mathbf{y}|\mathbf{c}_1)} \\
&\leq \sum_{\mathbf{y}:P(\mathbf{y}|\mathbf{c}_2)\geq P(\mathbf{y}|\mathbf{c}_1)} P(\mathbf{y}|\mathbf{c}_1)\frac{P(\mathbf{y}|\mathbf{c}_2)^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \\
&\leq \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{c}_1)\frac{P(\mathbf{y}|\mathbf{c}_2)^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \qquad \text{for any } s \geq 0 \\
&= \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{c}_1)^{1-s} P(\mathbf{y}|\mathbf{c}_2)^s
\end{aligned}$$

The choice $s = 1/2$ gives us

$$P_{e,1} \leq \sum_{\mathbf{y}} \sqrt{P(\mathbf{y}|\mathbf{c}_1)P(\mathbf{y}|\mathbf{c}_2)},$$

and by symmetry, the same quantity also upper bounds $P_{e,2}$.

For a memoryless channel $P(\mathbf{y}|\mathbf{c}_m) = \prod_{k=1}^n P(y_k|c_{m,k})$ and we obtain

$$\begin{aligned}
P_{e,m} &\leq \sum_{\mathbf{y}} \sqrt{P(\mathbf{y}|\mathbf{c}_1)P(\mathbf{y}|\mathbf{c}_2)} \\
&= \sum_{y_1} \cdots \sum_{y_n} \sqrt{P(y_1|c_{1,1})P(y_1|c_{2,1})} \ldots \sqrt{P(y_n|c_{1,n}P(y_n|c_{2,n})} \\
&= \Big[\sum_{y_1} \sqrt{P(y_1|c_{1,1})P(y_1|c_{2,1})}\Big] \ldots \Big[\sum_{y_n} \sqrt{P(y_n|c_{1,n})P(y_n|c_{2,n})}\Big] \\
&= \prod_{k=1}^{n} \Big[\sum_{y} \sqrt{P(y|c_{1,k})P(y|c_{2,k})}\Big].
\end{aligned}$$

For example, for a binary symmetric channel with crossover probability ϵ, we see that

$$\sum_{y} \sqrt{P(y|c_{1,k})P(y|c_{2,k})} = \begin{cases} 1 & \text{if } c_{1,k} = c_{2,k} \\ 2\sqrt{\epsilon(1-\epsilon)} & \text{else,} \end{cases}$$

and we obtain

$$P_{e,m} \leq [4\epsilon(1-\epsilon)]^{d/2}$$

where d is the number of places $\mathbf{c}_1$ and $\mathbf{c}_2$ are different (i.e., $d = |\{k : c_{1,k} \neq c_{2,k}\}|$).

4. Error Probability for two randomly chosen codewords

Suppose now that the codewords $\mathbf{c}_1$ and $\mathbf{c}_2$ are chosen randomly and independently each from a distribution Q on $\mathcal{X}^n$. Observe that the codewords are random variables $\mathbf{C}_1$ and

$\mathbf{C}_2$ and the probability that the block code is the particular block code with codewords $\mathbf{c}_1$ and $\mathbf{c}_2$ is given by $Q(\mathbf{c}_1)Q(\mathbf{c}_2)$. The error probabilities $P_{e,m}$ are now random variables $\mathbf{P}_{e,m}$ since they are functions of $\mathbf{C}_1$ and $\mathbf{C}_2$. Let $\bar{P}_{e,m}$ denote the expectation of $\mathbf{P}_{e,m}$.

We will now give an upper bound on $\bar{P}_{e,m}$ in two different ways. The first is the more straightforward, but the second way will turn out to be conceptually more useful. We will take $m = 1$, but is is clear by symmetry that $\bar{P}_{e,1} = \bar{P}_{e,2}$.

Method 1: Write

$$\begin{aligned}\bar{P}_{e,1} &= E[\mathbf{P}_{e,1}] \\ &= \sum_{\mathbf{c}_1}\sum_{\mathbf{c}_2} Q(\mathbf{c}_1)Q(\mathbf{c}_2)E[\mathbf{P}_{e,1}|\mathbf{C}_1 = \mathbf{c}_1, \mathbf{C}_2 = \mathbf{c}_2].\end{aligned}$$

But when we are given $\mathbf{c}_1$ and $\mathbf{c}_2$, $\mathbf{P}_{e,1}$ is no longer random, and we can use the bound of the last section on the probability of error to get

$$\begin{aligned}\bar{P}_{e,1} &\le \sum_{\mathbf{c}_1}\sum_{\mathbf{c}_2} Q(\mathbf{c}_1)Q(\mathbf{c}_2)\sum_{\mathbf{y}}\sqrt{P(\mathbf{y}|\mathbf{c}_1)}\sqrt{P(\mathbf{y}|\mathbf{c}_2)} \\ &= \sum_{\mathbf{y}}\Big[\sum_{\mathbf{c}_1} Q(\mathbf{c}_1)\sqrt{P(\mathbf{y}|\mathbf{c}_1)}\Big]\Big[\sum_{\mathbf{c}_2} Q(\mathbf{c}_2)\sqrt{P(\mathbf{y}|\mathbf{c}_2)}\Big] \\ &= \sum_{\mathbf{y}}\Big[\sum_{\mathbf{c}} Q(\mathbf{c})\sqrt{P(\mathbf{y}|\mathbf{c})}\Big]^2\end{aligned}$$

where the last equality follows by noting that the sums in the brackets in the next to last line are identical since they differ only by the index of summation.

Method 2: Write

$$\begin{aligned}\bar{P}_{e,1} &= E[\mathbf{P}_{e,1}] \\ &= \sum_{\mathbf{c}_1}\sum_{\mathbf{y}} Q(\mathbf{c}_1)P(\mathbf{y}|\mathbf{c}_1)E[\mathbf{P}_{e,1}|\mathbf{C}_1 = \mathbf{c}_1, \mathbf{Y} = \mathbf{y}].\end{aligned}$$

Note that here we are computing the expectation by first conditioning on the transmitted and received sequences. Now, observe that given $\mathbf{C}_1 = \mathbf{c}_1$ and the received sequence $\mathbf{y}$, an error is sure not to occur if $\mathbf{C}_2$ is chosen such that $P(\mathbf{y}|\mathbf{C}_2) < P(\mathbf{y}|\mathbf{c}_1)$, and otherwise we can upper bound $\mathbf{P}_{e,1}$ by 1. Thus, given $\mathbf{C}_1 = \mathbf{c}_1$ and $\mathbf{y}$, we have

$$\begin{aligned}\mathbf{P}_{e,1} &\le \begin{cases} 1 & \text{if } P(\mathbf{y}|\mathbf{C}_2) \ge P(\mathbf{y}|\mathbf{c}_1) \\ 0 & \text{if } P(\mathbf{y}|\mathbf{C}_2) < P(\mathbf{y}|\mathbf{c}_1) \end{cases} \\ &= \mathbb{1}_{P(\mathbf{y}|\mathbf{C}_2)\ge P(\mathbf{y}|\mathbf{c}_1)}.\end{aligned}$$

Taking expectations we see that

$$E[\mathbf{P}_{e,1}|\mathbf{C}_1 = \mathbf{c}_1, \mathbf{Y} = \mathbf{y}] \le \Pr(B_2)$$

where B_2 is the event that $\mathbf{C}_2$ is chosen such that $P(\mathbf{y}|\mathbf{C}_2) \ge P(\mathbf{y}|\mathbf{c}_1)$. We can bound $\Pr(B_2)$ by

$$\begin{aligned}\Pr(B_2) &= \sum_{\mathbf{c}_2} Q(\mathbf{c}_2)\mathbb{1}_{P(\mathbf{y}|\mathbf{c}_2)\ge P(\mathbf{y}|\mathbf{c}_1)} \\ &\le \sum_{\mathbf{c}_2} Q(\mathbf{c}_2)\frac{P(\mathbf{y}|\mathbf{c}_2)^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \qquad \text{for any } s \ge 0.\end{aligned}$$

Taking $s = 1/2$ and substituting back we obtain the same bound as before,

$$\bar{P}_{e,1} \le \sum_{\mathbf{y}} \Big[\sum_{\mathbf{c}} Q(\mathbf{c}) \sqrt{P(\mathbf{y}|\mathbf{c})} \Big]^2 .$$

For memoryless channels the bound simplifies if $Q(\mathbf{c})$ is chosen to be $Q(\mathbf{c}) = \prod_{k=1}^{n} Q(c_k)$. In this case we obtain

$$\bar{P}_{e,1} \le \Bigg[\sum_{y} \Big[\sum_{x} Q(x) \sqrt{P(y|x)} \Big]^2 \Bigg]^n .$$

5. Average error probability of a randomly chosen code

Consider now a code with M codewords, each codeword $\mathbf{c}_m$ chosen independently according to the probability distribution Q. Just as in the above section, the probability that the block code constructed is a particular block code with codewords $\mathbf{c}_1, \ldots, \mathbf{c}_M$ is $\prod_{m=1}^{M} Q(\mathbf{c}_m)$. The error probabilities $\mathbf{P}_{e,m}$ are also random variables, and again let $\bar{P}_{e,m}$ denote the expectation of $\mathbf{P}_{e,m}$. By the symmetry with respect to the permutation of the codewords in the construction, we see that $\bar{P}_{e,m}$ does not depend on m, and it will suffice to analyze $\bar{P}_{e,1}$. We will take the extension of 'method 2' in the previous section as our analysis method, and write

$$\bar{P}_{e,1} = \sum_{\mathbf{c}_1} \sum_{\mathbf{y}} Q(\mathbf{c}_1) P(\mathbf{y}|\mathbf{c}_1) E[\mathbf{P}_{e,1} | \mathbf{C}_1 = \mathbf{c}_1, \mathbf{Y} = \mathbf{y}].$$

Now, for a given $\mathbf{c}_1$ and $\mathbf{y}$ define for each $m \ge 2$, B_m as the event that codeword $\mathbf{C}_m$ is chosen such that $P(\mathbf{y}|\mathbf{C}_m) \ge P(\mathbf{y}|\mathbf{c}_1)$, i.e., that codeword m is at least as likely as the transmitted codeword. Then just as in the previous section,

$$\begin{aligned} E[\mathbf{P}_{e,1} | \mathbf{C}_1 = \mathbf{c}_1, \mathbf{Y} = \mathbf{y}] &\le \Pr\Big(\bigcup_{m=2}^{M} B_m \Big) \\ &\le \min\Big\{ 1, \sum_{m=2}^{M} \Pr(B_m) \Big\} \\ &\le \Big[\sum_{m=2}^{M} \Pr(B_m) \Big]^{\rho} \qquad \text{for all } \rho \in [0,1]. \end{aligned}$$

The second inequality above is just the union bound, the third inequality is because for $\rho \in [0,1]$, $x \le x^{\rho}$ when $x \in [0,1]$, and $1 \le x^{\rho}$ when $x \ge 1$.

Observe now that

$$\begin{aligned} \Pr(B_m) &= \sum_{\mathbf{c}_m} Q(\mathbf{c}_m) \mathbb{1}_{P(\mathbf{y}|\mathbf{c}_m) \ge P(\mathbf{y}|\mathbf{c}_1)} \\ &\le \sum_{\mathbf{c}_m} Q(\mathbf{c}_m) \frac{P(\mathbf{y}|\mathbf{c}_m)^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \qquad \text{for any } s \ge 0 \\ &= \sum_{\mathbf{c}} Q(\mathbf{c}) \frac{P(\mathbf{y}|\mathbf{c})^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \end{aligned}$$

and thus

$$E[\mathbf{P}_{e,1} | \mathbf{C}_1 = \mathbf{c}_1, \mathbf{Y} = \mathbf{y}] \le \Bigg[(M-1) \sum_{\mathbf{c}} Q(\mathbf{c}) \frac{P(\mathbf{y}|\mathbf{c})^s}{P(\mathbf{y}|\mathbf{c}_1)^s} \Bigg]^{\rho} .$$

Substituting back we get

$$\bar{P}_{e,1} \leq (M-1)^{\rho} \sum_{\mathbf{y}} \Big[\sum_{\mathbf{c}_1} Q(\mathbf{c}_1) P(\mathbf{y}|\mathbf{c}_1)^{1-s\rho} \Big] \Big[\sum_{\mathbf{c}} Q(\mathbf{c}) P(\mathbf{y}|\mathbf{c})^{s} \Big]^{\rho}.$$

Choosing now $s = 1/(1+\rho)$ (this choice in fact minimizes the bound) and observing that for this choice $1 - s\rho = s$, and that

$$\Big[\sum_{\mathbf{c}_1} Q(\mathbf{c}_1) P(\mathbf{y}|\mathbf{c}_1)^{1-s\rho} \Big] = \Big[\sum_{\mathbf{c}} Q(\mathbf{c}) P(\mathbf{y}|\mathbf{c})^{s} \Big]$$

since the two summations differ only by the summation index, we get

$$\bar{P}_{e,1} \leq (M-1)^{\rho} \sum_{\mathbf{y}} \Big[\sum_{\mathbf{c}} Q(\mathbf{c}) P(\mathbf{y}|\mathbf{c})^{1/(1+\rho)} \Big]^{1+\rho}.$$

If we now specialize this theorem to discrete memoryless channels and if we choose $Q(\mathbf{c}) = \prod_{i=1}^{n} Q(c_i)$, we get that for every $\rho \in [0,1]$,

$$\bar{P}_{e,1} \leq (M-1)^{\rho} \Bigg[\sum_{y} \Big[\sum_{x} Q(x) P(y|x)^{1/(1+\rho)} \Big]^{1+\rho} \Bigg]^{n}$$

If $M = \lceil e^{nR} \rceil$, then $(M-1) \leq e^{nR}$, and we can summarize the above as

THEOREM 1. *Given a discrete memoryless channel described by $P(y|x)$, for any blocklength n, and any $R \geq 0$ consider constructing a random block code with $M = \lceil e^{nR} \rceil$ codewords by choosing each letter of each codeword independently according to a distribution Q on $\mathcal{X}$. Then, the expected average error probability of this random code satisfies*

$$\bar{P}_{e,\text{ave}} \leq \exp\big\{ -n \max_{\rho \in [0,1]} [E_0(\rho, Q) - \rho R] \big\}$$

where

$$E_0(\rho, Q) = -\ln \sum_{y} \Big[\sum_{x} Q(x) P(y|x)^{1/(1+\rho)} \Big]^{1+\rho}.$$

Since the expected error probability cannot be better than the error probability of the best code we also get

COROLLARY 1. *Given a discrete memoryless channel described by $P(y|x)$, for any distribution Q on $\mathcal{X}$, any blocklength n and any $R > 0$, there exists a code of block length n and rate at least R with*

$$P_{e,\text{ave}} \leq \exp\big\{ -n E_r(R, Q) \big\}.$$

where

$$E_r(R, Q) = \max_{\rho \in [0,1]} [E_0(\rho, Q) - \rho R].$$

Note that the above corollary establishes the existence of codes of a certain rate with a guarantee on their $P_{e,\text{ave}}$, but suggests no mechanism to find them. However, if one does carry out the experiment of constructing a code by randomly choosing its codewords, the probability that the code obtained will be much worse than the average is small, in particular, Markov inequality tells us that the probability that a code constructed as such has $\mathbf{P}_{e,\text{ave}}$ larger than $\alpha \bar{P}_{e,\text{ave}}$ is small:

$$\Pr[\mathbf{P}_{e,\text{ave}} \geq \alpha \bar{P}_{e,\text{ave}}] \leq 1/\alpha \qquad \text{for } \alpha > 1.$$

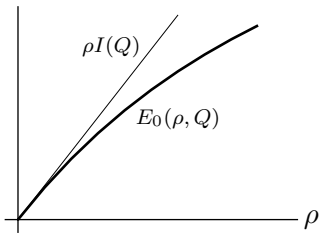

Figure 1: $E_0(\rho, Q)$

Thus, the difficulty in finding good practical codes is not that the good codes are rare, but the difficulty is in find good codes for which there are practical methods to encode and decode.

The corollary above makes guarantees on $P_{e,\text{ave}}$, but for a code constructed by random coding the value of $P_{e,\max}$ may be much higher than $P_{e,\text{ave}}$. Can we ensure the existence of codes with small $P_{e,\max}$?

THEOREM 2. *Given a discrete memoryless channel described by $P(y|x)$, for any distribution Q on $\mathcal{X}$, any blocklength n and any $R > 0$, there exists a block code of length n and rate at least R with*

$$P_{e,max} \leq 4 \exp\{-nE_r(R, Q)\}.$$

Proof. Let $M' = \lceil 2e^{nR} \rceil$. We know that there is a code with M' codewords and satisfies

$$P_{e,\text{ave}} = \frac{1}{M'} \sum_{m=1}^{M'} P_{e,m} \leq (M' - 1)^{\rho} \exp\{-nE_0(\rho, Q)\}$$

for every $\rho \in [0, 1]$. Since

$$(M' - 1)^{\rho} \leq 2^{\rho} e^{n\rho R} \leq 2e^{n\rho R},$$

we see that for this code

$$P_{e,\text{ave}} = \frac{1}{M'} \sum_{m=1}^{M'} P_{e,m} \leq 2 \exp\{-nE_r(R, Q)\}.$$

Now, among the M' numbers $P_{e,1}, \ldots, P_{e,M'}$ there cannot be more than $M'/2$ which exceed twice the average value $P_{e,\text{ave}}$. Thus, among these M' numbers there exist at least $\lceil e^{nR} \rceil$ of them which satisfy

$$P_{e,m} \leq 2P_{e,\text{ave}} \leq 4 \exp\{-nE_r(R, Q)\}.$$

Keeping only these codewords, and noticing that in the maximum likelihood decoder that corresponds to this code the decoding sets can only be enlarged, we see that we have constructed a code with the desired properties. □

5.1. PROPERTIES OF E_0 AND E_r

The value of the existence theorems we just proved depend on whether the bound we have on the probability of error can be made arbitrarily small for a set of rates of interest. Define

$$I(Q) = \sum_x \sum_y Q(x)P(y|x) \ln \frac{P(y|x)}{\sum_{x'} Q(x')P(y|x')}$$

as the value of mutual information between the input and output when the input distribution is Q. We will now show that for every rate $R < I(Q)$, the exponent $E_r(R, Q) > 0$, and thus for every rate $R < I(Q)$ we can find codes with arbitrarily small probability of error (by taking n large enough). Since this statement holds for every input distribution Q, it also follows that for every rate $R < \max_Q I(Q) = \max_{p_X} I(X;Y)$ there exist codes of arbitrarily small probability of error. This proves that any rate R below capacity is achievable (achievability part of the coding theorem).

A tedious but straightforward calculation also yields that

$$\left.\frac{\partial E_0(\rho, Q)}{\partial \rho}\right|_{\rho=0} = I(Q).$$

Using the fact that whenever Q and A are nonnegative, the function $t \in [0, \infty) \mapsto \left[\sum_x Q(x)A(x)^{1/t}\right]^t$ is nonincreasing, it is easy to show that $E_0(\rho, Q)$ is nondecreasing in ρ. We do not need it here, but it is also possible to show that $E_0(\rho, Q)$ is a concave function of ρ. Thus, Figure 1 shows the typical behavior of E_0 as a function of ρ.

Observe now that since the function $R \mapsto E_r(R, Q)$ is a maximum of the linear functions $R \mapsto E_0(\rho, Q) - \rho R$, (see Figure 2), we see that $E_r(R, Q) > 0$ whenever for some $\rho \in [0, 1]$, $E_0(\rho, Q) - \rho R > 0$ is satisfied.

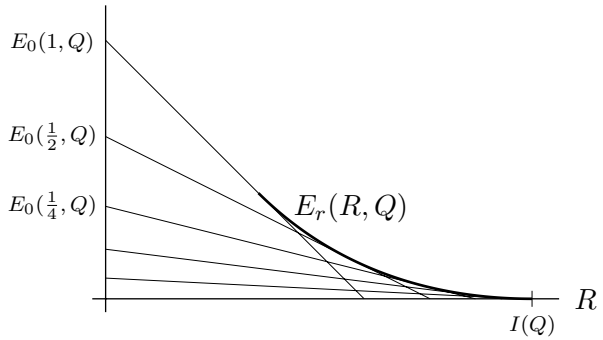

Figure 2: $E_r(R)$ as a maximum of linear functions.

It now follows that if $R < E_0(\rho, Q)/\rho$ for some $\rho \in (0, 1]$ and Q, then $E_0(\rho, Q) - \rho R > 0$ and thus $E_r(R, Q) > 0$. But,

$$I(Q) = \left.\frac{\partial E_0(\rho, Q)}{\partial \rho}\right|_{\rho=0} = \lim_{\rho \to 0} \frac{E_0(\rho, Q) - E_0(0, Q)}{\rho} = \lim_{\rho \to 0} \frac{E_0(\rho, Q)}{\rho}.$$

where the last equality follows from fact that $E_0(0, Q) = -\ln \sum_{x,y} Q(x)P(y|x) = -\ln 1 = 0$. Thus, by the definition of the limit above, for every $R < I(Q)$, we can find a ρ such that $E_0(\rho, Q)/\rho > R$. Hence for every $R < I(Q)$ we have $E_r(R, Q) > 0$. Since the above argument holds for any input distribution Q, it holds in particular for the Q that maximizes $I(Q)$. This proves the existence of codes with arbitrarily small probability of error for every rate less than the capacity $\max_Q I(Q)$.

The approach we used here yields to the function $E_r(R, Q)$ called *random coding error exponent*. To prove the achievability part of the coding theorem it would suffice to show that for any rate R below capacity there exists a sequence of rate R codes of length n with corresponding error probability that tends to zero as n tends to infinity. The random coding error exponent argument in addition to prove it, also characterizes the decay of the probability of error as n tends to infinity.