Chapter 8

Channel Capacity

1. *Preprocessing the output.* One is given a communication channel with transition probabilities $p(y \mid x)$ and channel capacity $C = \max_{p(x)} I(X;Y)$. A helpful statistician preprocesses the output by forming $\tilde{Y} = g(Y)$. He claims that this will strictly improve the capacity.

   (a) Show that he is wrong.

   (b) Under what conditions does he not strictly decrease the capacity?

   Solution: *Preprocessing the output.*

   (a) The statistician calculates $\tilde{Y} = g(Y)$. Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on x,

   $$I(X;Y) \geq I(X;\tilde{Y}). \tag{8.1}$$

   Let $\tilde{p}(x)$ be the distribution on x that maximizes $I(X;\tilde{Y})$. Then

   $$C = \max_{p(x)} I(X;Y) \geq I(X;Y)_{p(x)=\tilde{p}(x)} \geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} = \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}. \tag{8.2}$$

   Thus, the statistician is wrong and processing the output does not increase capacity.

   (b) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \to \tilde{Y} \to Y$ forming a Markov chain.

2. *Maximum likelihood decoding.* A source produces independent, equally probable symbols from an alphabet (a_1, a_2) at a rate of one symbol every 3 seconds. These symbols are transmitted over a binary symmetric channel which is used once each second by encoding the source symbol a_1 as 000 and the source symbol a_2 as 111. If in the corresponding 3 second interval of the channel output, any of the sequences 000,001,010,100 is received, a_1 is decoded; otherwise, a_2 is decoded. Let $\epsilon < \frac{1}{2}$ be the channel crossover probability.

(a) For each possible received 3-bit sequence in the interval corresponding to a given source letter, find the probability that a_1 came out of the source given that received sequence.

(b) Using part (a), show that the above decoding rule minimizes the probability of an incorrect decision.

(c) Find the probability of an incorrect decision (using part (a) is not the easy way here).

(d) If the source is slowed down to produce one letter every $2n+1$ seconds, a_1 being encoded by $2n+1$ 0's and a_2 being encoded by $2n+1$ 1's. What decision rule minimizes the probability of error at the decoder? Find the probability of error as $n \to \infty$.

Solution: *Maximum likelihood decoding.*

(a) By Bayes rule, for any events A and B,

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}. \tag{8.3}$$

In this case, we wish to calculate the conditional probability of a_1 given the channel output. Thus we take the event A to the event that the source produced a_1, and B to be the event corresponding to one of the 8 possible output sequences. Thus $\Pr(A) = 1/2$, and $\Pr(B|A) = \epsilon^i(1-\epsilon)^{3-i}$, where ϵ is the number of ones in the received sequence. $\Pr(B)$ can then be calculated as $\Pr(B) = \Pr(a_1)\Pr(B|a_1) + \Pr(B|a_2)$. Thus we can calculate

$$\Pr(a_1|000) = \frac{\frac{1}{2}(1-\epsilon)^3}{\frac{1}{2}(1-\epsilon)^3 + \frac{1}{2}\epsilon^3} \tag{8.4}$$

$$\Pr(a_1|100) = \Pr(a_1|010) = \Pr(a_1|001) = \frac{\frac{1}{2}(1-\epsilon)^2\epsilon}{\frac{1}{2}(1-\epsilon)^2\epsilon + \frac{1}{2}\epsilon^2(1-\epsilon)} \tag{8.5}$$

$$\Pr(a_1|110) = \Pr(a_1|011) = \Pr(a_1|101) = \frac{\frac{1}{2}(1-\epsilon)\epsilon^2}{\frac{1}{2}(1-\epsilon)\epsilon^2 + \frac{1}{2}\epsilon(1-\epsilon)^2} \tag{8.6}$$

$$\Pr(a_1|111) = \frac{\frac{1}{2}\epsilon^3}{\frac{1}{2}\epsilon^3 + \frac{1}{2}(1-\epsilon)^3} \tag{8.7}$$

(b) If $\epsilon < 1/2$, then the probability of a_1 given 000,001,010 or 100 is greater than 1/2, and the probability of a_2 given 110,011,101 or 111 is greater than 1/2. Therefore, the decoding rule above chooses the source symbol that has maximum probability given the observed output. This is the *maximum a posteriori* decoding rule, and is optimal in that it minimizes the probability of error. To see that this is true, let the input source symbol be X, let the output of the channel be denoted by Y and the decoded symbol be $\hat{X}(Y)$. Then

$$\begin{aligned}
\Pr(E) &= \Pr(X \neq \hat{X}) && (8.8)\\
&= \sum_y \Pr(Y=y)\Pr(X \neq \hat{X}|Y=y) && (8.9)\\
&= \sum_y \Pr(Y=y) \sum_{x \neq \hat{x}(y)} \Pr(x|Y=y) && (8.10)\\
&= \sum_y \Pr(Y=y)\left(1-\Pr(\hat{x}(y)|Y=y)\right) && (8.11)\\
&= \sum_y \Pr(Y=y) - \sum_y Pr(Y=y)\Pr(\hat{x}(y)|Y=y) && (8.12)\\
&= 1-\sum_y Pr(Y=y)\Pr(\hat{x}(y)|Y=y) && (8.13)
\end{aligned}$$

and thus to minimize the probability of error, we have to maximize the second term, which is maximized by choosing $\hat{x}(y)$ to the the symbol that maximizes the conditional probability of the source symbol given the output.

(c) The probability of error can also be expanded

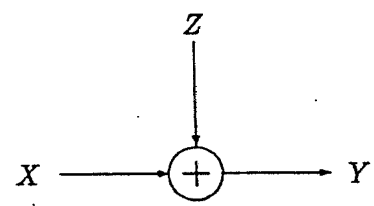
$$\begin{aligned}
\Pr(E) &= \Pr(X \neq \hat{X}) && (8.14)\\
&= \sum_x \Pr(x)\Pr(\hat{X} \neq x) && (8.15)\\
&= \Pr(a_1)\Pr(Y = 011, 110, 101, \text{ or } 111)\\
&\quad + \Pr(a_2)\Pr(Y = 000, 001, 010 \text{ or } 100) && (8.16)\\
&= \frac{1}{2}\left(3\epsilon^2(1-\epsilon)+\epsilon^3\right)+\frac{1}{2}\left(3\epsilon^2(1-\epsilon)+\epsilon^3\right) && (8.17)\\
&= 3\epsilon^2(1-\epsilon)+\epsilon^3. && (8.18)
\end{aligned}$$

(d) By extending the same arguments, it is easy to see that the decoding rule that minimizes the probability of error is the maximum a posteriori decoding rule, which in this case is the same as the maximum likelihood decoding rule (since the two input symbols are equally likely). So we choose the source symbol that is most likely to have produced the given output. This corresponds to choosing a_1 if the number of 1's in the received sequence is n or less, and choosing a_2 otherwise. The probability of error is then equal to (by symmetry) the probability of error given that a_1 was sent, which is the probability that $n+1$ or more 0's have been changed to 1's by the channel. This probability is

$$\Pr(E) = \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} \epsilon^i (1-\epsilon)^{2n+1-i} \qquad (8.19)$$

This probability goes to 0 as $n \to \infty$, since this is the probability that the number of 1's is $n+1$ or more, and since the expected proportion of 1's is $n\epsilon < n+1$, by the weak law of large numbers the above probability goes to 0 as $n \to \infty$.

3. *An additive noise channel.* Find the channel capacity of the following discrete memoryless channel:



where $\Pr\{Z=0\} = \Pr\{Z=a\} = \frac{1}{2}$. The alphabet for x is $\mathbf{X} = \{0,1\}$. Assume that Z is independent of X.

Observe that the channel capacity depends on the value of a.

Solution: *A sum channel.*

$$Y = X + Z \qquad X \in \{0,1\}, \quad Z \in \{0,a\} \tag{8.20}$$

We have to distinguish various cases depending on the values of a.

$a = 0$ In this case, $Y = X$, and $\max I(X;Y) = \max H(X) = 1$. Hence the capacity is 1 bit per transmission.

$a \neq 0, \pm 1$ In this case, Y has four possible values $0, 1, a$ and $1+a$. Knowing Y, we know the X which was sent, and hence $H(X|Y) = 0$. Hence $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

$a = 1$ In this case Y has three possible output values, $0, 1$ and 2, and the channel is identical to the binary erasure channel discussed in class, with $a = 1/2$. As derived in class, the capacity of this channel is $1 - a = 1/2$ bit per transmission.

$a = -1$ This is similar to the case when $a = 1$ and the capacity here is also $1/2$ bit per transmission.

4. *Channels with memory have higher capacity.* Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0,1\}$.

   Suppose that $\{Z_i\}$ has constant marginal probabilities $\Pr\{Z_i = 1\} = p = 1 - \Pr\{Z_i = 0\}$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Assume that Z^n is independent of the input X^n. Let $C = 1 - H(p, 1-p)$. Show that

$$\max_{p(x_1,x_2,\ldots,x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC.$$

Solution: *Channels with memory have a higher capacity.*

$$Y_i = X_i \oplus Z_i, \tag{8.21}$$

where

$$Z_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \tag{8.22}$$

and Z_i are not independent.

$$\begin{aligned} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \\ &= H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n | Y_1, Y_2, \ldots, Y_n) \\ &= H(X_1, X_2, \ldots, X_n) - H(Z_1, Z_2, \ldots, Z_n | Y_1, Y_2, \ldots, Y_n) \\ &\geq H(X_1, X_2, \ldots, X_n) - H(Z_1, Z_2, \ldots, Z_n) & (8.23) \\ &\geq H(X_1, X_2, \ldots, X_n) - \sum H(Z_i) & (8.24) \\ &= H(X_1, X_2, \ldots, X_n) - nH(p) & (8.25) \\ &= n - nH(p), & (8.26) \end{aligned}$$

if $X_1, X_2, \ldots, X_n$ are chosen i.i.d. $\sim$ Bern$(\frac{1}{2})$. The capacity of the channel with memory over n uses of the channel is

$$\begin{aligned} nC^{(n)} &= \max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) & (8.27) \\ &\geq I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n)_{p(x_1, x_2, \ldots, x_n) = \text{Bern}(\frac{1}{2})} & (8.28) \\ &\geq n(1 - H(p)) & (8.29) \\ &= nC. & (8.30) \end{aligned}$$

Hence channels with memory have higher capacity. The intuitive explanation for this result is that the correlation between the noise decreases the effective noise; one could use the information from the past samples of the noise to combat the present noise.

5. *Channel capacity.* Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ 1/3, & 1/3, & 1/3 \end{pmatrix}$$

and $X \in \{0, 1, \ldots, 10\}$. Assume that Z is independent of X.

(a) Find the capacity.

(b) What is the maximizing $p^*(x)$?

Solution: *Channel capacity.*

$$Y = X + Z (\text{mod } 11) \tag{8.31}$$

where

$$Z = \begin{cases} 1 & \text{with probability} 1/3 \\ 2 & \text{with probability} 1/3 \\ 3 & \text{with probability} 1/3 \end{cases}. \tag{8.32}$$

In this case,

$$H(Y|X) = H(Z|X) = H(Z) = \log 3, \tag{8.33}$$

independent of the distribution of X, and hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) && (8.34)\\ &= \max_{p(x)} H(Y) - H(Y|X) && (8.35)\\ &= \max_{p(x)} H(Y) - \log 3 && (8.36)\\ &= \log 11 - \log 3, && (8.37) \end{aligned}$$

which is attained when Y has an uniform distribution, which occurs (by symmetry) when X has an uniform distribution.

(a) The capacity of the channel is $\log \frac{11}{3}$ bits/transmission.

(b) The capacity is achieved by an uniform distribution on the inputs. $p(X = i) = \frac{1}{11}$ for $i = 0, 1, \ldots, 10$.

6. *Using two channels at once.* Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1 \mid x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2 \mid x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1 \mid x_1) \times p(y_2 \mid x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of this channel.

Solution: *Using two channels at once.* Suppose we are given two channels, $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$, which we can use at the same time. We can define the product channel as the channel, $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$. To find the capacity of the product channel, we must find the distribution $p(x_1, x_2)$ on the input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ that maximizes $I(X_1, X_2; Y_1, Y_2)$. Since the joint distribution

$$p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2), \tag{8.38}$$

$Y_1 \to X_1 \to X_2 \to Y_2$ forms a Markov chain and therefore

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) && (8.39)\\ &= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) && (8.40)\\ &= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (8.41)\\ &\le H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) && (8.42)\\ &= I(X_1; Y_1) + I(X_2; Y_2), && (8.43) \end{aligned}$$

where (8.40) and (8.41) follow from Markovity, and we have equality in (8.42) if Y_1 and Y_2 are independent. Equality occurs when X_1 and X_2 are independent. Hence

$$\begin{aligned} C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) && (8.44)\\ &\le \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) && (8.45)\\ &= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) && (8.46)\\ &= C_1 + C_2. && (8.47) \end{aligned}$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions that maximize C_1 and C_2 respectively.

7. *Noisy typewriter.* Consider a 26-key typewriter.

   (a) If pushing a key results in printing the associated letter, what is the capacity C in bits?

   (b) Now suppose that pushing a key results in printing that letter or the next (with equal probability)? Thus $A \to A$ or $B, \ldots, Z \to Z$ or A. What is the capacity?

   (c) What is the highest rate code with block length one that you can find that achieves *zero* probability of error for the channel in part (b) .

Solution: *Noisy typewriter.*

(a) If the typewriter prints out whatever key is struck, then the output, Y, is the same as the input, X, and

$$C = \max I(X;Y) = \max H(X) = \log 26, \tag{8.48}$$

attained by a uniform distribution over the letters.

(b) In this case, the output is either equal to the input (with probability $\frac{1}{2}$) or equal to the next letter (with probability $\frac{1}{2}$). Hence $H(Y|X) = \log 2$ independent of the distribution of X, and hence

$$C = \max I(X;Y) = \max H(Y) - \log 2 = \log 26 - \log 2 = \log 13, \tag{8.49}$$

attained for a uniform distribution over the output, which in turn is attained by a uniform distribution on the input.

(c) A simple zero error block length one code is the one that uses every alternate letter, say A,C,E,... ,W,Y. In this case, none of the codewords will be confused, since A will produce either A or B, C will produce C or D, etc. The rate of this code,

$$R = \frac{\log(\#\text{ codewords})}{\text{Block length}} = \frac{\log 13}{1} = \log 13. \tag{8.50}$$

In this case, we can achieve capacity with a simple code with zero error.

8. *Cascade of Binary Symmetric Channels.* Show that a cascade of n identical binary symmetric channels,

$$X_0 \to \boxed{\text{BSC \#1}} \to X_1 \to \cdots \to X_{n-1} \to \boxed{\text{BSC \#n}} \to X_n$$

each with raw error probability p, is equivalent to a single BSC with error probability $\frac{1}{2}(1-(1-2p)^n)$ and hence that $\lim_{n\to\infty} I(X_0;X_n) = 0$ if $p \neq 0, 1$. No encoding or decoding takes place at the intermediate terminals $X_1, \ldots, X_{n-1}$. Thus the capacity of the cascade tends to zero.

Solution: *Cascade of binary symmetric channels.* There are many ways to solve this problem. One way is to use the singular value decomposition of the transition probability matrix for a single BSC.

Let,

$$A = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

be the transition probability matrix for our BSC. Then the transition probability matrix for the cascade of n of these BSC's is given by,

$$A_n = A^n.$$

Now check that,

$$A = T^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1-2p \end{bmatrix} T$$

where,

$$T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Using this we have,

$$\begin{aligned} A_n &= A^n \\ &= T^{-1} \begin{bmatrix} 1 & 0 \\ 0 & (1-2p)^n \end{bmatrix} T \\ &= \begin{bmatrix} \frac{1}{2}(1+(1-2p)^n) & \frac{1}{2}(1-(1-2p)^n) \\ \frac{1}{2}(1-(1-2p)^n) & \frac{1}{2}(1+(1-2p)^n) \end{bmatrix}. \end{aligned}$$

From this we see that the cascade of n BSC's is also a BSC with probablility of error,

$$p_n = \frac{1}{2}(1-(1-2p)^n).$$

The matrix, T, is simply the matrix of eigenvectors of A.

This problem can also be solved by induction on n.

Probably the simplest way to solve the problem is to note that the probability of error for the cascade channel is simply the sum of the odd terms of the binomial expansion of $(x+y)^n$ with $x=p$ and $y=1-p$. But this can simply be written as $\frac{1}{2}(x+y)^n - \frac{1}{2}(y-x)^n = \frac{1}{2}(1-(1-2p)^n$.

9. *The Z channel.* The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0, 1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

Solution: *The Z channel.* First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $x = \Pr(X=1)$:

$$\begin{aligned} H(Y|X) &= \Pr(X=0)\cdot 0 + \Pr(X=1)\cdot 1 = x \\ H(Y) &= H(\Pr(Y=1)) = H(x/2) \\ I(X;Y) &= H(Y) - H(Y|X) = H(x/2) - x \end{aligned}$$

Since $I(X;Y) = 0$ when $x = 0$ and $x = 1$, the maximum mutual information is obtained for some value of x such that $0 < x < 1$.

Using elementary calculus, we determine that

$$\frac{d}{dx} I(X;Y) = \frac{1}{2}\log_2 \frac{1-x/2}{x/2} - 1,$$

which is equal to zero for $x = 2/5$. (It is reasonable that $\Pr(X=1) < 1/2$ because $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel in bits is $H(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.

10. *Suboptimal codes.* For the Z channel of the previous problem, assume that we choose a $(2^{nR}, n)$ code at random, where each codeword is a sequence of *fair* coin tosses. This will not achieve capacity. Find the maximum rate R such that the probability of error $P_e^{(n)}$, averaged over the randomly generated codes, tends to zero as the block length n tends to infinity.

    Solution: *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

$$I(X;Y) = H(Y) - H(Y|X) = H(\frac{3}{4},\frac{1}{4}) - \frac{1}{2}H(\frac{1}{2},\frac{1}{2}) = \frac{3}{2} - \frac{3}{4}\log 3. \tag{8.51}$$

11. *Zero-error capacity.* A channel with alphabet $\{0,1,2,3,4\}$ has transition probabilities of the form

$$p(y|x) = \begin{cases} 1/2 & \text{if } y = x \pm 1 \bmod 5 \\ 0 & \text{otherwise.} \end{cases}$$

    (a) Compute the capacity of this channel in bits.

    (b) The zero-error capacity of a channel is the number of bits per channel use that can be transmitted with zero probability of error. Clearly, the zero-error capacity of this pentagonal channel is at least 1 bit (transmit 0 or 1 with probability 1/2). Find a block code that shows that the zero-error capacity is greater than 1 bit. Can you estimate the exact value of the zero-error capacity?
    (*Hint*: Consider codes of length 2 for this channel.)
    The zero-error capacity of this channel was found by Lovasz[8].

Solution: *Zero-error capacity.*

(a) Since the channel is symmetric, it is easy to compute its capacity:

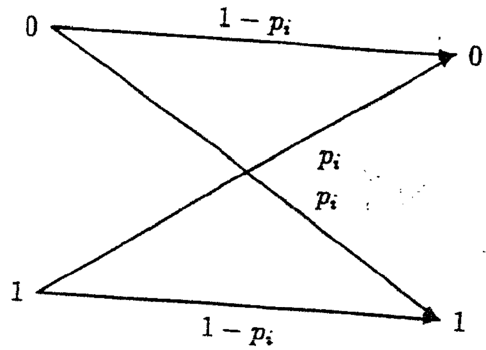
$$H(Y|X) = 1$$
$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - 1.$$

So mutual information is maximized when Y is uniformly distributed, which occurs when the input X is uniformly distributed. Therefore the capacity in bits is $C = \log_2 5 - 1 = \log_2 2.5 = 1.32$.

(b) Let us construct a block code consisting of 2-tuples. We need more than 4 codewords in order to achieve capacity greater than 2 bits, so we will pick 5 codewords with distinct first symbols: $\{0a, 1b, 2c, 3d, 4e\}$. We must choose a, b, c, d, e so that the receiver will be able to determine which codeword was transmitted. A simple repetition code will not work, since if, say, 22 is transmitted, then 11 might be received, and the receiver could not tell whether the codeword was 00 or 22. Instead, we use the code $\{03, 14, 20, 31, 42\}$; that is, each codeword is of the form uv, where $v = u + 3 \bmod 5$. Then whenever xy is received, there is exactly one possible codeword. (Each codeword will be received as one of 4 possible 2-tuples; so there are 20 possible received 2-tuples, out of a total of 25.) Since there are 5 possible error-free messages with 2 channel uses, the zero-error capacity of this channel is at least $\frac{1}{2}\log_2 5 = 1.161$ bits.

In fact, the zero-error capacity of this channel is exactly $\frac{1}{2}\log_2 5$. This result was obtained by László Lovász, "On the Shannon capacity of a graph," *IEEE Transactions on Information Theory*, Vol IT-25, pp. 1–7, January 1979. The first results on zero-error capacity are due to Claude E. Shannon, "The zero-error capacity of a noisy channel," *IEEE Transactions on Information Theory*, Vol IT-2, pp. 8–19, September 1956, reprinted in *Key Papers in the Development of Information Theory*, David Slepian, editor, IEEE Press, 1974.

12. *Time-varying channels.* Consider a time-varying discrete *memoryless* channel. Let $Y_1, Y_2, \ldots, Y_n$ be conditionally independent given $X_1, X_2, \ldots, X_n$, with conditional distribution given by $p(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{n} p_i(y_i \mid x_i)$.



Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$. Find $\max_{p(\mathbf{x})} I(\mathbf{X}; \mathbf{Y})$.

Solution: *Time-varying channels.*

We can use the same chain of inequalities as in the proof of the converse to the channel coding theorem. Hence

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n) \tag{8.52}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n) \tag{8.53}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i), \tag{8.54}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything else. Continuing the series of inequalities, we have

$$I(X^n;Y^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \tag{8.55}$$

$$\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \tag{8.56}$$

$$\le \sum_{i=1}^{n} (1 - h(p_i)), \tag{8.57}$$

with equality if $X_1, X_2, \ldots, X_n$ is chosen i.i.d. $\sim$ Bern(1/2). Hence

$$\max_{p(\mathbf{x})} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n} (1 - h(p_i)). \tag{8.58}$$