17 — Communication over Constrained Noiseless Channels

Solution to exercise 17.10 (p.255). Let the invariant distribution be

\[ P(s) = \alpha e_s^{(L)} e_s^{(R)}, \]  

(17.22)

where \( \alpha \) is a normalization constant. The entropy of \( S_t \) given \( S_{t-1} \), assuming \( S_{t-1} \) comes from the invariant distribution, is

\[ H(S_t|S_{t-1}) = - \sum_{s, s'} P(s) P(s'|s) \log P(s'|s) \]

(17.23)

\[ = - \sum_{s, s'} \alpha e_s^{(L)} e_s^{(R)} \frac{e_s^{(L)} A_{s's}}{\lambda e_s^{(L)}} \log \frac{e_s^{(L)} A_{s's}}{\lambda e_s^{(L)}} \]

(17.24)

\[ = - \sum_{s, s'} \alpha e_s^{(R)} e_s^{(L)} \frac{A_{s's}}{\lambda} \left[ \log e_s^{(L)} + \log A_{s's} - \log \lambda - \log e_s^{(L)} \right]. \]

(17.25)

Now, \( A_{s's} \) is either 0 or 1, so the contributions from the terms proportional to \( A_{s's} \) are all zero. So

\[ H(S_t|S_{t-1}) = \log \lambda + \frac{\alpha}{\lambda} \sum_s \left( \sum_s A_{s's} e_s^{(R)} \right) e_s^{(L)} \log e_s^{(L)} + \]

\[ \frac{\alpha}{\lambda} \sum_s \left( \sum_s e_s^{(L)} A_{s's} \right) e_s^{(R)} \log e_s^{(L)} \]

(17.26)

\[ = \log \lambda - \frac{\alpha}{\lambda} \sum_s \lambda e_s^{(R)} e_s^{(L)} \log e_s^{(L)} + \frac{\alpha}{\lambda} \sum_s \lambda e_s^{(L)} e_s^{(R)} \log e_s^{(L)} \]

(17.27)

\[ = \log \lambda. \]

(17.28)

Solution to exercise 17.11 (p.255). The principal eigenvalues of the connection matrices of the two channels are 1.839 and 1.928. The capacities (\( \log \lambda \)) are 0.879 and 0.947 bits.

Solution to exercise 17.12 (p.256). The channel is similar to the unconstrained binary channel; runs of length greater than \( L \) are rare if \( L \) is large, so we only expect weak differences from this channel; these differences will show up in contexts where the run length is close to \( L \). The capacity of the channel is very close to one bit.

A lower bound on the capacity is obtained by considering the simple variable-length code for this channel which replaces occurrences of the maximum runlength string 111...1 by 111...0, and otherwise leaves the source file unchanged. The average rate of this code is \( 1/(1+2^{-L}) \) because the invariant distribution will hit the ‘add an extra zero’ state a fraction \( 2^{-L} \) of the time.

We can reuse the solution for the variable-length channel in exercise 6.18 (p.125). The capacity is the value of \( \beta \) such that the equation

\[ Z(\beta) = \sum_{l=1}^{L+1} 2^{-\beta l} = 1 \]

(17.29)

is satisfied. The \( L+1 \) terms in the sum correspond to the \( L+1 \) possible strings that can be emitted, 0, 10, 110, …, 11...10. The sum is exactly given by:

\[ Z(\beta) = 2^{-\beta} \frac{(2^{-\beta})^{L+1} - 1}{2^{-\beta} - 1}. \]

(17.30)
17.7: Solutions

Here we used \[ \sum_{n=0}^{N} ar^n = \frac{a(r^{N+1} - 1)}{r - 1}. \]

We anticipate that \( \beta \) should be a little less than 1 in order for \( Z(\beta) \) to equal 1. Rearranging and solving approximately for \( \beta \), using \( \ln(1 + x) \approx x \),

\[
Z(\beta) = 1 \\
\Rightarrow \beta \approx 1 - 2^{-(L+2)/\ln 2}.
\]  (17.31) (17.32)

We evaluated the true capacities for \( L = 2 \) and \( L = 3 \) in an earlier exercise. The table compares the approximate capacity \( \beta \) with the true capacity for a selection of values of \( L \).

<table>
<thead>
<tr>
<th>( L )</th>
<th>( \beta )</th>
<th>True capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.910</td>
<td>0.879</td>
</tr>
<tr>
<td>3</td>
<td>0.955</td>
<td>0.947</td>
</tr>
<tr>
<td>4</td>
<td>0.977</td>
<td>0.975</td>
</tr>
<tr>
<td>5</td>
<td>0.9887</td>
<td>0.9881</td>
</tr>
<tr>
<td>6</td>
<td>0.9944</td>
<td>0.9942</td>
</tr>
<tr>
<td>9</td>
<td>0.9993</td>
<td>0.9993</td>
</tr>
</tbody>
</table>

The element \( Q_{10} \) will be close to 1/2 (just a tiny bit larger), since in the unconstrained binary channel \( Q_{10} = 1/2 \). When a run of length \( L - 1 \) has occurred, we effectively have a choice of printing 10 or 0. Let the probability of selecting 10 be \( f \). Let us estimate the entropy of the remaining \( N \) characters in the stream as a function of \( f \), assuming the rest of the matrix \( Q \) to have been set to its optimal value. The entropy of the next \( N \) characters in the stream is the entropy of the first bit, \( H_2(f) \), plus the entropy of the remaining characters, which is roughly \( (N-1) \) bits if we select 0 as the first bit and \( (N-2) \) bits if 1 is selected. More precisely, if \( C \) is the capacity of the channel (which is roughly 1),

\[
H(\text{the next } N \text{ chars}) \approx H_2(f) + [(N-1)(1-f) + (N-2)f] C \\
= H_2(f) + NC - fC \approx H_2(f) + N - f. \]  (17.33)

Differentiating and setting to zero to find the optimal \( f \), we obtain:

\[
\log_2 \frac{1-f}{f} \approx 1 \Rightarrow \frac{1-f}{f} \approx 2 \Rightarrow f \approx 1/3. \]  (17.34)

The probability of emitting a 1 thus decreases from about 0.5 to about 1/3 as the number of emitted 1s increases.

Here is the optimal matrix:

\[
\begin{bmatrix}
0 & 0.3334 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.4287 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.4669 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.4841 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.4923 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.4963 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4983 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4993 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0.6667 & 0.5713 & 0.5334 & 0.5159 & 0.5077 & 0.5037 & 0.5017 & 0.5007 & 0.5002
\end{bmatrix}.
\]  (17.35)

Our rough theory works.