

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

$$P(s) = \alpha e_s^{(L)} e_s^{(R)}, \quad (17.22)$$

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

Here, as in Chapter 4, S_t denotes the ensemble whose random variable is the state s_t .

$$H(S_t|S_{t-1}) = - \sum_{s,s'} P(s)P(s'|s) \log P(s'|s) \quad (17.23)$$

$$= - \sum_{s,s'} \alpha e_s^{(L)} e_s^{(R)} \frac{e_{s'}^{(L)} A_{s's}}{\lambda e_s^{(L)}} \log \frac{e_{s'}^{(L)} A_{s's}}{\lambda e_s^{(L)}} \quad (17.24)$$

$$= - \sum_{s,s'} \alpha e_s^{(R)} \frac{e_{s'}^{(L)} A_{s's}}{\lambda} \left[\log e_{s'}^{(L)} + \log A_{s's} - \log \lambda - \log e_s^{(L)} \right]. \quad (17.25)$$

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

$$H(S_t|S_{t-1}) = \log \lambda + -\frac{\alpha}{\lambda} \sum_{s'} \left(\sum_s A_{s's} e_s^{(R)} \right) e_{s'}^{(L)} \log e_{s'}^{(L)} + \frac{\alpha}{\lambda} \sum_s \left(\sum_{s'} e_{s'}^{(L)} A_{s's} \right) e_s^{(R)} \log e_s^{(L)} \quad (17.26)$$

$$= \log \lambda - \frac{\alpha}{\lambda} \sum_{s'} \lambda e_{s'}^{(R)} e_{s'}^{(L)} \log e_{s'}^{(L)} + \frac{\alpha}{\lambda} \sum_s \lambda e_s^{(L)} e_s^{(R)} \log e_s^{(L)} \quad (17.27)$$

$$= \log \lambda. \quad (17.28)$$

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 $111\dots 1$ by $111\dots 10$, 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 β such that the equation

$$Z(\beta) = \sum_{l=1}^{L+1} 2^{-\beta l} = 1 \quad (17.29)$$

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

$$Z(\beta) = 2^{-\beta} \frac{(2^{-\beta})^{L+1} - 1}{2^{-\beta} - 1}. \quad (17.30)$$

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

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

$$Z(\beta) = 1 \tag{17.31}$$

$$\Rightarrow \beta \simeq 1 - 2^{-(L+2)} / \ln 2. \tag{17.32}$$

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

L	β	True capacity
2	0.910	0.879
3	0.955	0.947
4	0.977	0.975
5	0.9887	0.9881
6	0.9944	0.9942
9	0.9993	0.9993

The element $Q_{1|0}$ will be close to $1/2$ (just a tiny bit larger), since in the unconstrained binary channel $Q_{1|0} = 1/2$. 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 \mathbf{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),

$$\begin{aligned} H(\text{the next } N \text{ chars}) &\simeq H_2(f) + [(N - 1)(1 - f) + (N - 2)f] C \\ &= H_2(f) + NC - fC \simeq H_2(f) + N - f. \end{aligned} \tag{17.33}$$

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

$$\log_2 \frac{1-f}{f} \simeq 1 \Rightarrow \frac{1-f}{f} \simeq 2 \Rightarrow f \simeq 1/3. \tag{17.34}$$

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:

$$\begin{bmatrix} 0.3334 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .4287 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .4669 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .4841 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .4923 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .4963 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4983 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4993 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4998 \\ 1.6666 & .5713 & .5331 & .5159 & .5077 & .5037 & .5017 & .5007 & .5002 \end{bmatrix}. \tag{17.35}$$

Our rough theory works.