Solutions: Homework Set # 4

Problem 1

(a) The stationary distribution is:

$$\mu_0 = \frac{p_{10}}{p_{01}+p_{10}}, \quad \mu_0 = \frac{p_{01}}{p_{01}+p_{10}}$$

So the entropy rate is

$$H(X_2|X_1) = \mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{10}H(p_{01}) + p_{01}H(p_{10})}{p_{01}+p_{10}}$$

(b) Since the process has only two states the entropy rate is at most 1 bit and it can be achieved iff $p_{01} = p_{10} = 1/2$.

(c) The entropy rate is

$$H(X_2|X_1) = \mu_0 H(p) + \mu_1 H(1) = \frac{H(p)}{p+1}$$

(d) After some mathematics,we see that the maximum happens at $p^\star = (3-\sqrt{5})/2 = 0.382$ and the maximum will be:

$$\max_p H(X_2|X_1) = \frac{H(p^\star)}{1+p^\star} = 0.694 \qquad \text{bits}$$

Note that $\frac{1}{p^\star}$ is the Golden Ratio!

(e) Note that the Markov chain of part (c) doesn't allow consecutive ones. Consider any allowable sequence of symbols of length t. If the first symbol is 1, then the next symbol must be 0. The remaining $N(t-2)$ remaining symbols can make any allowable sequence. If the first symbol is 0, then the remaining $N(t-1)$ symbols can make any allowable sequences. So we have:

$$N(t) = N(t-1) + N(t-2) \qquad N(1) = 2, N(2) = 3$$

The sequence $N(t)$ grows exponentially, i.e. $N(t) \approx c\lambda^t$ where λ is te solution of the characteristic equation

$$1 = z^{-1} + z^{-2}$$

Solving this equation yields to $\lambda = (1+\sqrt{5})/2$. Therefore

$$h_0 = \lim_{n\to\infty} \frac{1}{t} \log N(t) = \log \frac{1+\sqrt{5}}{2} = 0.694 \qquad \text{bits}$$

Since there are only $N(t)$ possible outcomes for $X_1, X_2, \cdots, X_t$ an upper bound on $H(X_1, X_2, \cdots, X_t)$ is $\log N(t)$, so the entropy rate of the Markov chain of part (c) is at most H_0. And we saw in part (d) that this bound is achievable.

Problem 2

(a) $\log(n+1)$ bits are needed to be reserved for the description of k.

$$N = \binom{n}{k}$$

(b) We need $\lceil \log \binom{n}{k} \rceil$ bits to describe which of the N sequences we are decoding with.

(c)

$$\begin{aligned} l(x^n) &= \lceil \log(n+1) \rceil + \lceil \log \binom{n}{k} \rceil \\ &\leq \log(n) + \log \binom{n}{k} + 3 \end{aligned} \tag{1}$$

Since $\binom{n}{k} \leq 2^{nH(\frac{k}{n})} \sqrt{\frac{n}{\pi k(n-k)}}$

$$\begin{aligned} \log \binom{n}{k} &\leq \log \left(2^{nH(\frac{k}{n})} \sqrt{\frac{n}{\pi k(n-k)}}\right) \\ &= nH(\frac{k}{n}) + \log \sqrt{\frac{n}{\pi k(n-k)}} \end{aligned} \tag{2}$$

So

$$l(x^n) \leq \log n + nH(\frac{k}{n}) - \frac{1}{2}\log n + \log \sqrt{\frac{n^2}{\pi k(n-k)}} + 3 \tag{3}$$

$$= \frac{1}{2}\log n + nH(\frac{k}{n}) \underbrace{-\frac{1}{2}\log \left(\pi \frac{k}{n} \times \frac{n-k}{n}\right) + 3}_{\text{does not grow with } n} \tag{4}$$

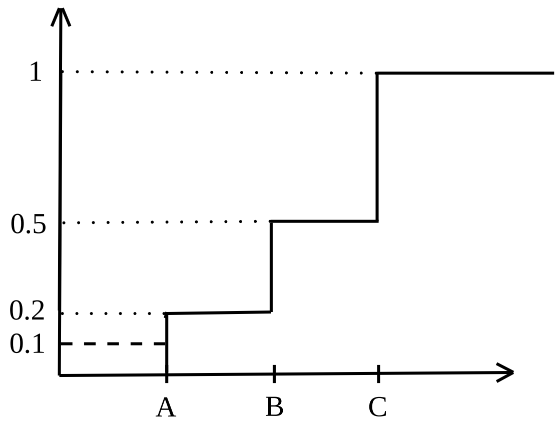
(d)

$$\frac{l(x^n) - l^*(x^n)}{l^*(x^n)} \leq \frac{\frac{1}{2}\log n + nH(\frac{k}{n}) + O(c) - nH(\frac{k}{n})}{nH(\frac{k}{n})} = \frac{1/2 \log n}{nH(\frac{k}{n})} + \frac{O(c)}{nH(\frac{k}{n})} \tag{5}$$

$$\lim_{n \to \infty} \frac{l(x^n) - l^*(x^n)}{l^*(x^n)} = 0 \tag{6}$$

Problem 3

(a) Initially, the input sequence is empty and the corresponding interval is $[0, 1)$. The CDF after the first input symbol is:



Thus, the interval corresponding to A is $[0, 0.2)$.

(b) When the second symbol (C) arrives, the interval corresponding to (AC) is $[0.1, 0.2)$.

(c) To encode the midpoint of these intervals Shannon-Fano-Eliad coding is used.

So, the binary representation of A is the binary representation of 0.1, which is 0.00011001....

As discussed in class, $\lceil \log(\frac{1}{0.2}) \rceil + 1 = 4$ is the length of the codeword assigned to the point 0.1, thus A is encoded to 0001

The interval corresponding to AC is $[0.1, 0.2)$ and the midpoint is 0.15 (0.0010011...in binary). The length of this description is $\lceil \log(\frac{1}{0.1}) \rceil + 1 = 5$, thus the representation is 00100.

(d) $ACCB$ has a corresponding interval $[0.16, 0.175)$ and the probability of the sequence is $0.2 \cdot 0.5 \cdot 0.5 \cdot 0.3 = 0.1 \cdot 0.15 = 0.015$.

The midpoint is 0.167 which is 0.00101010101.... The length of description is $\lceil \log(\frac{1}{0.015}) \rceil + 1 = 8$. Thus, the codeword representing $ACCB$ is 00101010.

(e) If the sequence to be encoded is represented by the interval $[a, b)$, the bits of the binary representation of $\frac{a+b}{2}$ which we can be sure of (i.e. they are fixed no matter how the sequence continues) are the common bits of the binary representations of a and b. This is because based on how the sequence continues, the interval can converge to any point in the interval (extremes: a and b).

So, to answer (e) we know that $a = 0.16$ and $b = 0.175$.

$$(a)_2 = 0.001010\ldots \quad (7)$$

$$(b)_2 = 0.001011\ldots \quad (8)$$

Thus, 5 bits is what we can be sure of (out of the 8 bit representation).

Problem 4

(a) The parsing is as follows:

$$0, 00, 000, 1, 10, 101, 0000, 01, 1010, 1 \ldots \tag{9}$$

and the encoding is

$$(0000, 0), (0001, 0), (0010, 0), (0000, 1), (0100, 0), (0101, 1), (0011, 0), (0001, 1), (0110, 0), \ldots \tag{10}$$

(b)

$$0, 1, 00, 01, 10, 11, \ldots \tag{11}$$

(c)

$$0, 00, 000, 0000, \ldots \tag{12}$$