Homework 2 solutions

1. Problem 2.48 on pg. 54 of the text.

Solution:

Sequence length

(a) Since \( N \) is a deterministic function of \( X^N \), we have \( I(N; X^N) = H(N) - H(N | X^N) = H(N) \). Since \( P(N = k) = 2^{-k} \) for \( k \geq 1 \), we have

\[
H(N) = \sum_{k=1}^{\infty} -2^{-k} \log 2^{-k} = \sum_{k=1}^{\infty} k2^{-k} = 2.
\]

(b) Since \( X^N \) is a deterministic function of \( N \), we have \( H(X^N | N) = 0 \).

(c) We have

\[
H(X^N) = I(X^N; N) + H(X^N | N) = I(X^N; N) = 2.
\]

• Note that in the remaining three parts of the problem we are dealing with a different pair of random variables \((N, X^N)\) than in the first three parts.

(d) Since \( N \) is a deterministic function of \( X^N \) we have

\[
I(N; X^N) = H(N) - H(N | X^N) = H(N) = \frac{1}{3} \log 3 + \frac{2}{3} \log \frac{3}{2} = \log 3 - \frac{2}{3}.
\]

(e) Conditioned on \( N = 6 \), \( X^N \) is uniformly distributed on a set of size \( 2^6 \), while, conditioned on \( N = 12 \), \( X^N \) is uniformly distributed on a set of size \( 2^{12} \). Thus

\[
H(X^N | N) = H(X^N | N = 6)P(N = 6) + H(X^N | N = 12)P(N = 12) = \frac{1}{3}6 + \frac{2}{3}12 = 10.
\]

(f) We have

\[
H(X^N) = I(N; X^N) + H(X^N | N) = \log 3 - \frac{2}{3} + 10 = \log 3 + \frac{28}{3}.
\]

2. Problem 3.3 on pg. 65 of the text.

Solution:

Piece of cake

Let \( X_k = 1 \) if the \( k \)-th cut of the cake is done in the proportions \((\frac{2}{3}, \frac{1}{3})\), and let \( X_k = 0 \) if it is done in the proportions \((\frac{2}{5}, \frac{3}{5})\). Then \( X_k, k \geq 1 \) are i.i.d. with \( P(X_k = 1) = \frac{2}{3} \) and \( P(X_k = 0) = \frac{1}{3} \). The size of the piece of cake after \( n \) cuts is

\[
\prod_{k=1}^{n} \left( \frac{2}{3} \right)^{X_k} \left( \frac{3}{5} \right)^{1-X_k} = \left( \frac{3}{5} \right)^n \prod_{k=1}^{n} \left( \frac{10}{9} \right)^{X_k}.
\]
The logarithm of the size of the cake after \( n \) cuts is therefore

\[
n(\log \frac{3}{5} + (\frac{1}{n} \sum_{k=1}^{n} X_k) \log \frac{10}{9}) .
\]

By the weak law of large numbers, for every \( \epsilon > 0 \) we have

\[
P(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - \frac{3}{4} \right| > \epsilon) \to 0 \text{ as } n \to \infty .
\]

Hence, for every \( \epsilon > 0 \),

\[
P(\left| \log \frac{3}{5} + (\frac{1}{n} \sum_{k=1}^{n} X_k) \log \frac{10}{9} \right| - \left[ \log \frac{3}{5} + \frac{3}{4} \log \frac{10}{9} \right] > \epsilon) \to 0 \text{ as } n \to \infty .
\]

In this sense, we can say that the size of piece of cake after \( n \) cuts, to the first order in the exponent, is \( \log \frac{3}{5} + \frac{3}{4} \log \frac{10}{9} \).

3. Problem 3.12 on pg. 68 of the text.

Solution:

Monotonic convergence of the empirical distribution;

(a) Let \( \hat{p}_n' \) be the empirical distribution of the samples \( n + 1 \) through \( 2n \), i.e.

\[
\hat{p}_n'(x) = \frac{1}{n} \sum_{i=1}^{n} 1(X_{n+i} = x) \text{ for all } x \in \mathcal{X} .
\]

Then we have \( \hat{p}_{2n} = \frac{1}{2} (\hat{p}_n + \hat{p}_n') \). By the convexity of relative entropy, we have

\[
D(\hat{p}_{2n} \| p) \leq \frac{1}{2} \left( D(\hat{p}_n \| p) + D(\hat{p}_n' \| p) \right) .
\]

Note that in this expression all three empirical distributions \( \hat{p}_{2n} \), \( \hat{p}_n \), and \( \hat{p}_n' \) are random variables taking values in the space of probability distributions on \( \mathcal{X} \), so this equation is really a pointwise identity between nonnegative real valued random variables. Taking expectations on both sides of this equation gives

\[
E[D(\hat{p}_{2n} \| p)] \leq \frac{1}{2} \left( E[D(\hat{p}_n \| p)] + E[D(\hat{p}_n' \| p)] \right) = E[D(\hat{p}_n \| p)] ,
\]

where the last step is justified because \( E[D(\hat{p}_n' \| p)] = E[D(\hat{p}_n \| p)] \). This is what was to be shown.

(b) For \( 1 \leq i \leq n \), let \( \hat{p}_n^{(i)} \) denote the empirical distribution of the \( n - 1 \) random variables \( X_j, 1 \leq j \leq n, j \neq i \), i.e.

\[
\hat{p}_n^{(i)}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} 1(X_j = x) \text{ for all } x \in \mathcal{X} .
\]
Then we have
\[ \hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{p}^{(i)}_n. \]

Using the convexity of the relative entropy function, this yields
\[ D(\hat{p}_n \| p) \leq \frac{1}{n} \sum_{i=1}^{n} D(\hat{p}^{(i)}_n \| p), \]
which is a pointwise equality between nonnegative real valued random variables. Taking expectations on both sides gives
\[ E[D(\hat{p}_n \| p)] \leq \frac{1}{n} \sum_{i=1}^{n} E[D(\hat{p}^{(i)}_n \| p)] = E[D(\hat{p}_{n-1} \| p)], \]
where the last step is justified because \( E[D(\hat{p}^{(i)}_n \| p)] = E[D(\hat{p}_{n-1} \| p)] \) for all \( i \).

4. Problem 4.2 on pp. 88 -89 of the text.

Solution:

\textit{Time’s arrow}

We have
\[ H(X_0 \mid X_{-1}, \ldots, X_{-n}) = H(X_{-n}, \ldots, X_0) - H(X_{-n}, \ldots, X_{-1}). \]

We also have
\[ H(X_0 \mid X_1, \ldots, X_n) = H(X_0, \ldots, X_n) - H(X_1, \ldots, X_n). \]

By the stationarity of the process, we have
\[ H(X_{-n}, \ldots, X_0) = H(X_0, \ldots, X_n) \]
and
\[ H(X_{-n}, \ldots, X_{-1}) = H(X_1, \ldots, X_n), \]
from which the claim follows.

5. Problem 4.7 on pg. 92 of the text.

Solution:

\textit{Entropy rates of Markov chains}

(a) The stationary distribution of the chain is given by
\[ \pi_0 = \frac{p_{10}}{p_{01} + p_{10}}, \quad \pi_1 = \frac{p_{01}}{p_{01} + p_{10}}. \quad (1) \]

The entropy rate is therefore
\[ H(X) = \frac{p_{10}}{p_{01} + p_{10}} h(p_{01}) + \frac{p_{01}}{p_{01} + p_{10}} h(p_{10}). \quad (2) \]
where $h(p)$ is the binary entropy function.

(b) Since $h(p) \leq 1$ for all $p \in [0, 1]$, we see from (2) that $H(X) \leq 1$. But equality is achieved if $p_{01} = p_{10} = \frac{1}{2}$. Thus these are the values which maximize the entropy rate, and the maximum entropy rate is 1.

(c) The stationary distribution of the chain is given by

$$
\pi_0 = \frac{1}{1 + p}, \quad \pi_1 = \frac{p}{1 + p}.
$$

The entropy rate is therefore $\frac{h(p)}{1 + p}$.

(d) The value of $p$ that maximizes this entropy rate can be found by differentiating. We get $\frac{3 - \sqrt{5}}{2}$ for this value. Note that this is less than $\frac{1}{2}$. The corresponding maximum entropy rate is $\log \frac{2}{\sqrt{5} - 1}$.

(e) Let $N(t, 0)$, (resp. $N(t, 1)$) denote the number of allowable state sequence $s$ of length $t$ ending in state 0 (resp. state 1). Then $N(t) = N(t, 0) + N(t, 1)$. Also,

$$
N(t, 0) = N(t-1, 0) + N(t-1, 1) = N(t-1)
$$

$$
N(t, 1) = N(t-1, 0) = N(t-2, 0) + N(t-2, 1) = N(t-2)
$$

so that we have $N(t) = N(t-1) + N(t-2)$ for $t \geq 3$. This is the recursion for the well-known Fibonacci sequence. Here the initial values are $N(1) = 2$ and $N(2) = 3$. (This differs from what is usually called the Fibonacci sequence, which has initial conditions 1 and 1.) The sequence can be described by its generating function $n(z) = \sum_{t=1}^{\infty} N(t) z^t$, thought of as a function on the complex plane. Using the recursion gives

$$
n(z) = \frac{z^2 + 2z}{1 - z - z^2}.
$$

The zeros of the denominator are at $\frac{1 \pm \sqrt{5}}{2}$. The smaller of these in absolute value is $\frac{1 - \sqrt{5}}{2}$, so the radius of convergence of $n(z)$ around the origin is $\frac{\sqrt{5} - 1}{2}$. From this we get

$$
\lim_{t \to \infty} \log \frac{N(t)}{t} = \log \frac{2}{\sqrt{5} - 1}.
$$

Since the uniform distribution on a finite set maximizes entropy, we have $H(X_1, \ldots, X_t) \leq \log N(t)$ for any choice of $p$ in part (c). It follows that the right hand side of (5) is an upper bound for the entropy rate of the Markov chain in part (c) for any choice of $p$. Note that for the choice $p = \frac{3 - \sqrt{5}}{2}$, the entropy rate $\log \frac{2}{\sqrt{5} - 1}$ calculated in part (d) equals the upper bound given by the right hand side of (5).

This is an instance of a general phenomenon: the number of paths in a time invariant trellis grows exponentially in the number of stages of the trellis and the exponent is given to first order by the maximum entropy rate among all Markov chains whose state transition diagram is compatible with a single stage of the trellis. The subject of symbolic dynamics is concerned with related questions (about sequences of paths with restricted transitions) and has had several significant applications, e.g. to many commercially deployed codes for magnetic recording.
6. Problem 4.19 on pp. 96 -97 of the text.

Solution:

Random walk on graph

(a) The stationary distribution of the Markov chain on the vertices of the graph is

\[
\begin{bmatrix}
\frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{4}{16}
\end{bmatrix}.
\]

(b) Let \((X_n, -\infty < n < \infty)\) denote the stationary process. Since the process is a Markov chain on the vertices of the graph its entropy rate equals \(H(X_1 \mid X_0)\). This is \(4 \cdot \frac{3}{16} \log 3 + \frac{4}{16} \log 4\).

(c) For the stationary process, we have

\[
I(X_{n+1}; X_n) = H(X_{n+1}) - H(X_{n+1} \mid X_n) = \left[4 \cdot \frac{3}{16} \log \frac{16}{3} + \frac{4}{16} \log \frac{16}{4}\right] - \left[4 \cdot \frac{3}{16} \log 3 + \frac{4}{16} \log 4\right],
\]

which simplifies to

\[
I(X_{n+1}; X_n) = \frac{3}{4} \log \frac{16}{9}.
\]

7. Problem 4.26 on pg. 98 of the text.

Solution:

Transitions in Markov chains

Each of \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_{n-1})\) is a deterministic function of the other (we are thinking of the one sided process starting at time 1). Thus

\[
H(X_1, \ldots, X_n) = H(Y_1, \ldots, Y_{n-1}).
\]

Hence we get

\[
\lim_{n \to \infty} \frac{1}{n-1} H(Y_1, \ldots, Y_{n-1}) = \lim_{n \to \infty} \frac{n}{n-1} \frac{1}{n} H(X_1, \ldots, X_n),
\]

which tell us that the edge process has the same entropy rate as the original process. It should be pointed out that the proof has nothing to do with the original process being Markov. Also, the one sided nature of the process definitions is not significant, since in the two-sided case we would have

\[
\left|\frac{1}{n-1} H(Y_1, \ldots, Y_n) - \frac{1}{n-1} H(X_1, \ldots, X_n)\right| = \frac{1}{n-1} H(X_0 \mid X_1, \ldots, X_n) \to 0 \text{ as } n \to \infty.
\]