EL606: Homework 3 – Solutions

Due date: Oct. 5, 2004

1. *Problem 4, Ch. 4, Cover&Thomas*
   Let X_k denote $X_1, X_2, \ldots X_k$.

   (a) By the chain rule for entropy,

   $$\frac{H(X^n)}{n} = \frac{H(X_n|X^{n-1}) + H(X^{n-1})}{n}. \tag{1}$$

   From stationarity it follows that for all $1 \leq i \leq n$,

   $$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

   which further implies, by averaging both sides, that

   $$H(X_n|X^{n-1}) \leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} = \frac{H(X^{n-1})}{n-1}. \tag{2}$$

   Combining (1) and (2) yields,

   $$\frac{H(X^n)}{n} \leq \frac{1}{n}\left[\frac{H(X^{n-1})}{n-1} + H(X^{n-1})\right] = \frac{H(X^{n-1})}{n-1}.$$

   (b) By stationarity we have for all $1 \leq i \leq n$,

   $$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

   which further implies, by averaging both sides, that

   $$H(X_n|X^{n-1}) = \frac{\sum_{i=1}^{n} H(X_n|X^{n-1})}{n} \leq \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} = \frac{H(X_1, X_2, \ldots, X_n)}{n}.$$

2. *Problem 5, Ch. 4, Cover&Thomas*

   (a) The stationary distribution is (see the text)

   $$\mu_0 = \frac{p_{10}}{p_{01} + p_{10}} \text{ and } \mu_1 = \frac{p_{01}}{p_{01} + p_{10}}.$$

   Therefore 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) The entropy rate is at most 1 bit because the process has only two states. This rate can be achieved if (and only if) $p_{01} = p_{10} = 1/2$, in which case the process is actually i.i.d. with $\Pr(X_i = 0) = \Pr(X_i = 1) = 1/2$.

   (c) As a special case of the general two-state Markov chain, the entropy rate is

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

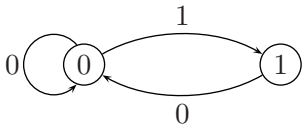


Figure 1: Graph for the Golden Ratio Markov Chain.

(d) By straightforward calculus, we find that the maximum value of $H(X)$ of part (c) occurs for $p = (3-\sqrt{5})/2 = 0.382$. The maximum value is

$$H(p) = H(1-p) = H\left(\frac{\sqrt{5}-1}{2}\right) = 0.694 \text{ bits}\,.$$

Note that $(\sqrt{5}-1)/2 = 0.618$ is (the reciprocal of) the Golden Ratio.

(e) The Markov chain of part (c) forbids consecutive ones as shown in Fig. 1. It is easy to see that the following recursion holds:

$$\begin{aligned} &N(1) = 2,\ N(2) = 3, \\ &N(t) = N(t-1) + N(t-2),\ n \geq 3. \end{aligned} \tag{3}$$

The sequence $N(t)$ grows exponentially, that is, $N(t) \approx c\lambda^t$, where λ is the maximum magnitude solution of the characteristic equation

$$1 = z^{-1} + z^{-2}\,.$$

Solving the characteristic equation yields $\lambda = (1+\sqrt{5})/2$, the Golden Ratio. Therefore

$$H_0 = \lim_{n\to\infty} \frac{1}{t} \log N(t) = \log(1+\sqrt{5})/2 = 0.694 \text{ bits}\,.$$

Since there are only $N(t)$ possible outcomes for $X_1, \ldots, X_t$, an upper bound on the entropy $H(X_1, \ldots, X_t)$ is $\log N(t)$, and so the entropy rate of the Markov chain of part (c) is at most H_0. We saw in part (d) that this upper bound can be achieved by setting $p = (3-\sqrt{5})/2 = 0.382$.

3. *Problem 8, Ch. 4, Cover&Thomas*
We first prove, by induction, that for any $k \leq n-1$, the probability that $S_k = \sum_{i=1}^{k} X_i$ is odd is $1/2$. Clearly this is true for $k = 1$. Assume that it is true for $k-1$. Then

$$\begin{aligned} \boxed{P(S_k \text{ odd})} &= P(S_{k-1} \text{ odd})P(X_k = 0) + P(S_{k-1} \text{ even})P(X_k = 1) \\ &= \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \boxed{\frac{1}{2}}. \end{aligned}$$

Hence,

$$P(X_n = 1) = P(X_n = 0) = \frac{1}{2}.$$

(a) Since $X_1, \ldots, X_{n-1}$ are i.i.d, when i and j are both less than n, X_i and X_j are independent. The only possible problem is when $j = n$.

$$\begin{aligned}\boxed{P(X_i = 1, X_n = 1)} &= P(X_i = 1, \sum_{k=1, k\neq i}^{n-1} X_k \text{ even}) \\ &= P(X_i = 1) \cdot P\Big(\sum_{k=1, k\neq i}^{n-1} X_k \text{ even}\Big) \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \boxed{P(X_i = 1) \cdot P(X_n = 1)}\end{aligned}$$

Therefore, X_i and X_n are independent.

(b) Since X_i and X_j are independent and uniformly distributed on $\{0, 1\}$,

$$H(X_i, X_j) = H(X_i) + H(X_j) = 1 + 1 = 2 \text{ bits.}$$

(c) By the chain rule and the independence of $X_1, X_2, \ldots, X_{n_1}$, we have

$$\begin{aligned}H(X_1, X_2, \ldots, X_n) &= H(X_1, X_2, \ldots, X_{n-1}) + H(X_n | X_{n-1}, \ldots, X_1) \\ &= \sum_{i=1}^{n-1} H(X_i) + 0 \\ &= n - 1,\end{aligned}$$

since X_n is a function of the previous X_i's. The total entropy is not n, which is what would be obtained if the X_i's were all independent. This example illustrates that pairwise independence does not imply complete independence.

4. *Problem 9, Ch. 4, Cover&Thomas*

(a) $H(X_n|X_0) = H(X_{-n}|X_0)$.
This statement is true, since

$$\begin{aligned}H(X_n|X_0) &= H(X_n, X_0) - H(X_0) \\ H(X_{-n}|X_0) &= H(X_{-n}, X_0) - H(X_0)\end{aligned}$$

and $H(X_n, X_0) = H(X_{-n}, X_0)$ by stationarity.

(b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.
This statement is not true in general, though it is true for first order Markov chains. A simple counterexample is a periodic process with period n. Let $X_0, X_1, X_2, \ldots, X_{n-1}$ be i.i.d. uniformly distributed binary random variables and let $X_k = X_{k-n}$ for $k \geq n$. In this case, $H(X_n|X_0) = 0$ and $H(X_{n-1}|X_0) = 1$, contradicting the statement $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

(c) $H(X_n|X_1^{n-1}, X_{n+1})$ is non-increasing in n.
This statement is true, since by stationarity $H(X_n|X_1^{n-1}, X_{n+1}) = H(X_{n+1}|X_2^n, X_{n+2}) \geq H(X_{n+1}|X_1^n, X_{n+2})$ where the inequality follows from the fact that conditioning reduces entropy.