EE 376A/Stat 376A
Information Theory
Prof. T. Cover

Handout #13
Tuesday, Feburary 3, 2009

Solutions to Homework Set #3

1. **Random walk in a cube.**
   A bird flies from room to room in a $3 \times 3 \times 3$ cube (equally likely through each interior wall). What is the entropy rate?

   Solution: Random walk in a cube.
   The entropy rate of a random walk on a graph with equal weights is given by equation 4.41 in the text:
   $$H(\mathcal{X}) = \log(2E) - H\left(\frac{E_1}{2E}, \ldots, \frac{E_m}{2E}\right)$$
   There are 8 corners, 12 edges, 6 faces and 1 center. Corners have 3 edges, edges have 4 edges, faces have 5 edges and centers have 6 edges. Therefore, the total number of edges $E = 54$. So,
   $$\begin{aligned} H(\mathcal{X}) =& \log(108) + 8\left(\frac{3}{108}\log\frac{3}{108}\right) + 12\left(\frac{4}{108}\log\frac{4}{108}\right) \\ &+ 6\left(\frac{5}{108}\log\frac{5}{108}\right) + 1\left(\frac{6}{108}\log\frac{6}{108}\right) \\ =& 2.03 \text{ bits.} \end{aligned}$$

2. **Entropy of graphs.**
   Consider a random walk on a (connected) graph with 3 edges.
   (a) Which graph has the lowest entropy rate? What is the rate?
   (b) Which has the highest entropy rate?

   Solution: Entropy of graphs.
   There are three choices for graphs with 3 edges (See Figure 1).
   The entropy rate is given by:
   $$\begin{aligned} \mathcal{H} &= -\sum_i \mu_i \sum_j P_{ij} \log P_{ij} \\ &= \sum_i \frac{W_i}{W} \log(W_i) \end{aligned}$$

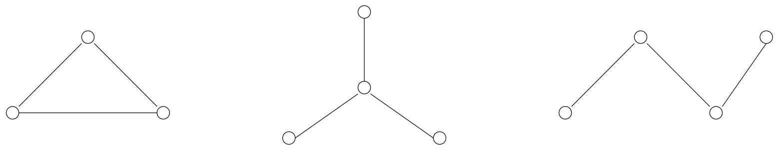

Figure 1: Graphs with three edges

(a) Graph 1
$\{W_i\} = \{2,2,2\} \;\Rightarrow\; \mathcal{H} = 3\big(\frac{2}{6}\log(2)\big) \;=\; 1$

(b) Graph 2
$\{W_i\} = \{1,1,1,3\} \;\Rightarrow\; \mathcal{H} = \frac{3}{6}\log(3) \;=\; 0.79$

(c) Graph 3
$\{W_i\} = \{1,2,2,1\} \;\Rightarrow\; \mathcal{H} = 2\big(\frac{2}{6}\log(2)\big) \;=\; 0.667$

Thus, Graph 1 has highest entropy while Graph 3 has the lowest.

3. **Stationary processes.**
Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be a stationary (not necessarily Markov) stochastic process. Which of the following statements are true? Prove or provide a counterexample.

(a) $H(X_n|X_0) = H(X_{-n}|X_0)$.

(b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

(c) $H(X_n|X_1^{n-1}, X_{n+1})$ is nonincreasing in n.

(d) $H(X_{n+1}|X_1^n, X_{n+2}^{2n+1})$ is nonincreasing in n.

Solution: Stationary processes.

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

$$H(X_n|X_0) = H(X_n, X_0) - H(X_0) \tag{1}$$
$$H(X_{-n}|X_0) = H(X_{-n}, X_0) - H(X_0) \tag{2}$$

and $H(X_n, X_0) = H(X_{-n}, X_0)$ by stationarity. (Note that $\Pr(X_n = a|X_0 = b) \neq \Pr(X_0 = a|X_n = b)$ in general.)

(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. Bern($\frac{1}{2}$) random variables and let $X_{mN+k} = X_k$ for $k = 0, \ldots, N-1$ and $m = 1, -1, 2, -2, \ldots$. Note that this is a stationary process. In this case, for $n = mN$, $H(X_n|X_0) = 0$ 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.

(d) $H(X_{n+1}|X_1^n, X_{n+2}^{2n+1})$ is non-increasing in n.

This statement is true, since by stationarity,

$$H(X_{n+1}|X_1^n, X_{n+2}^{2n+1}) = H(X_{n+2}|X_2^{n+1}, X_{n+3}^{2n+2}) \geq H(X_{n+2}|X_1^{n+1}, X_{n+3}^{2n+3}),$$

where the inequality follows from the fact that conditioning reduces entropy.

4. **Entropy Rate of Second-order Markov.**

Let $\{X_n\}$, $X_n \in \{0, 1\}$ be a binary stochastic process. Let $X_{n+1} = X_n \oplus X_{n-1} \oplus Z_{n+1}$, where $\oplus$ is addition modulo 2 and $\{Z_n\}$ is a Bernoulli(p) process. What is the entropy rate of $\{X_n\}$? It may be useful to redefine this second order Markov process as a first order Markov process.

Solution: Entropy Rate of Second order Markov.

Let $Y_i = (X_i, X_{i-1})$ for i. Then it is easy to see that $\lim_{n\to\infty} H(X^n)/n = \lim_{n\to\infty} H(Y^n)/n$, and $\{Y_i\}$ is a first order Markov process. Hence the entropy rate of $\{X_n\}$ is the same as the entropy rate of the process $\{Y_n\}$. Since $\{Y_n\}$ is a first order Markov process, we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} H(Y^n) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y_{i-1}) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(p) \\ &= H(p). \end{aligned}$$

We conclude that $\{X_n\}$ has an entropy rate of $H(p)$.

Alternatively, we can directly attack the entropy rate of the second order Markov process $\{X_n\}$ as

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}H(X^n) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(X_i|X^{i-1}) \\
&\overset{(a)}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(X_i|X_{i-1},X_{i-2}) \\
&\overset{(b)}{\to} H(p),
\end{aligned}$$

where (a) follows from the second-order Markovity of X_n and (b) follows from the fact that $H(X_n|X_{n-1},X_{n-2}) = H(p)$ for all $n \geq 3$ (and from Cesáro mean theorem).

5. **The past has little to say about the future.**
For a stationary stochastic process $X_1, X_2, \ldots$, show that

$$\lim_{n\to\infty}\frac{1}{2n}I(X_1,X_2,\ldots,X_n;X_{n+1},X_{n+2},\ldots,X_{2n}) = 0.$$

Thus the dependence between adjacent n-blocks of a stationary process does not grow linearly with n.

Solution: The past has little to say about the future.
From stationarity, it is easy to see that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{2n}I(X_1^n;X_{n+1}^{2n}) &= \lim_{n\to\infty}\frac{1}{2n}\big(H(X_1^n)+H(X_{n+1}^{2n})-H(X_1^{2n})\big) \\
&= \lim_{n\to\infty}\frac{1}{2n}\big(2H(X_1^n)-H(X_1^{2n})\big) \\
&= \lim_{n\to\infty}\left(\frac{1}{n}H(X_1^n)-\frac{1}{2n}H(X_1^{2n})\right) \\
&= \left(\lim_{n\to\infty}\frac{1}{n}H(X_1^n)\right)-\left(\lim_{n\to\infty}\frac{1}{2n}H(X_1^{2n})\right) \\
&= H(\mathcal{X})-H(\mathcal{X}) \\
&= 0
\end{aligned}$$

The second-to-last step holds because we know that for a stationary stochastic process, the entropy rate exists and is equal to $\lim_{n\to\infty}\frac{1}{n}H(X_1^n)$.

6. **Markov chain.**

$$P = [P_{ij}] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Let X_1 be uniformly distributed over the states $\{0, 1, 2\}$. Let $\{X_i\}_1^\infty$ be a Markov chain with transition matrix P, thus $P(X_{n+1} = j | X_n = i) = P_{ij}, i, j \in \{0, 1, 2\}$.

(a) Is $\{X_n\}$ stationary?

(b) Find $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$.

Now consider the derived process $Z_1, Z_2, \ldots, Z_n$, where

$$\begin{aligned} Z_1 &= X_1 \\ Z_i &= X_i - X_{i-1} \pmod 3, \qquad i = 2, \ldots, n. \end{aligned}$$

Thus Z^n encodes the transitions, not the states.

(c) Find $H(Z_1, Z_2, ..., Z_n)$.

(d) Find $H(Z_n)$ and $H(X_n)$, for $n \geq 2$.

(e) Find $H(Z_n | Z_{n-1})$ for $n \geq 2$.

(f) Are Z_{n-1} and Z_n independent for $n \geq 2$?

Solution: Markov chain.

(a) Let μ_n denote the probability mass function at time n. Since $\mu_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\mu_2 = \mu_1 P = \mu_1$, $\mu_n = \mu_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for all n and $\{X_n\}$ is stationary.
Alternatively, the observation P is doubly stochastic will lead the same conclusion.

(b) Since $\{X_n\}$ is stationary Markov,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) &= H(X_2 | X_1) \\ &= \sum_{k=0}^{2} P(X_1 = k) H(X_2 | X_1 = k) \\ &= 3 \times \frac{1}{3} \times H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) \\ &= \frac{3}{2}. \end{aligned}$$

(c) Since $(X_1, \ldots, X_n)$ and $(Z_1, \ldots, Z_n)$ are one-to-one, by the chain rule of entropy and the Markovity,

$$\begin{aligned}
H(Z_1,\dots,Z_n) &= H(X_1,\dots,X_n) \\
&= \sum_{k=1}^{n} H(X_k|X_1,\dots,X_{k-1}) \\
&= H(X_1) + \sum_{k=2}^{n} H(X_k|X_{k-1}) \\
&= H(X_1) + (n-1)H(X_2|X_1) \\
&= \log 3 + \frac{3}{2}(n-1).
\end{aligned}$$

Alternatively, we can use the results of parts (d), (e), and (f). Since $Z_1,\dots,Z_n$ are independent and $Z_2,\dots,Z_n$ are identically distributed with the probability distribution $(\frac{1}{2},\frac{1}{4},\frac{1}{4})$,

$$\begin{aligned}
H(Z_1,\dots,Z_n) &= H(Z_1) + H(Z_2) + \dots + H(Z_n) \\
&= H(Z_1) + (n-1)H(Z_2) \\
&= \log 3 + \frac{3}{2}(n-1).
\end{aligned}$$

(d) Since $\{X_n\}$ is stationary with $\mu_n = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$,

$$H(X_n) = H(X_1) = H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) = \log 3.$$

For $n \geq 2$, $Z_n = \begin{cases} 0, & \frac{1}{2}, \\ 1, & \frac{1}{4}, \\ 2, & \frac{1}{4}. \end{cases}$

Hence, $H(Z_n) = H(\frac{1}{2},\frac{1}{4},\frac{1}{4}) = \frac{3}{2}$.

(e) Due to the symmetry of P, $P(Z_n|Z_{n-1}) = P(Z_n)$ for $n \geq 2$. Hence, $H(Z_n|Z_{n-1}) = H(Z_n) = \frac{3}{2}$.

Alternatively, using the result of part (f), we can trivially reach the same conclusion.

(f) Let $k \geq 2$. First observe that by the symmetry of P, $Z_{k+1} = X_{k+1} - X_k$ is independent of X_k. Now that

$$\begin{aligned}
P(Z_{k+1}|X_k, X_{k-1}) &= P(X_{k+1} - X_k|X_k, X_{k-1}) \\
&= P(X_{k+1} - X_k|X_k) \\
&= P(X_{k+1} - X_k) \\
&= P(Z_{k+1}),
\end{aligned}$$

Z_{k+1} is independent of (X_k, X_{k-1}) and hence independent of $Z_k = X_k - X_{k-1}$.

For $k = 1$, again by the symmetry of P, Z_2 is independent of $Z_1 = X_1$ trivially.

7. **Waiting times.**
Let X be the waiting time for the first heads to appear in successive flips of a fair coin. For example, $\Pr\{X = 3\} = (\frac{1}{2})^3$. Let S_n be the waiting time for the nth head to appear. Thus,

$$S_0 = 0$$
$$S_{n+1} = S_n + X_{n+1}$$

where $X_1, X_2, X_3, \ldots$ are i.i.d according to the distribution above.

(a) Is the process $\{S_n\}$ stationary?

(b) Calculate $H(S_1, S_2, \ldots, S_n)$.

(c) Does the process $\{S_n\}$ have an entropy rate? If so, what is it? If not, why not?

(d) What is the expected number of fair coin flips required to generate a random variable having the same distribution as S_n?

Solution: Waiting times.

(a) For the process to be stationary, the distribution must be time invariant. It turns out that process $\{S_n\}$ is not stationary. There are several ways to show this.

  i. S_0 is always 0 while S_i, $i \neq 0$ can take on several values. Since the marginals for S_0 and S_1, for example, are not the same, the process can't be stationary.

  ii. It's clear that the variance of S_n grows with n, which again implies that the marginals are not time-invariant.

  iii. Process $\{S_n\}$ is an independent increment process. It is a standard result from ee278 and related probability-theory classes that an independent increment process is not stationary (not even wide sense stationary!).

(b) We can use chain rule and Markov properties to obtain the following results.

$$\begin{aligned}
H(S_1, S_2, \ldots, S_n) &= H(S_1) + \sum_{i=2}^{n} H(S_i|S^{i-1}) \\
&= H(S_1) + \sum_{i=2}^{n} H(S_i|S_{i-1}) \\
&= H(X_1) + \sum_{i=2}^{n} H(X_i) \\
&= \sum_{i=1}^{n} H(X_i) \\
&= 2n
\end{aligned}$$

(c) If your answer in part (b) is correct, then part (c) should be trivial.

$$\begin{aligned}\mathcal{H}(S) &= \lim_{n\to\infty} \frac{H(S^n)}{n} \\ &= \lim_{n\to\infty} \frac{2n}{n} \\ &= 2\end{aligned}$$

Note that the entropy rate can still exist even when the process is not stationary. Furthermore, the entropy rate (for this problem) is the same as the entropy of X.

(d) It turned out that this problem was harder than the others. Many students answered $2n$ which turned out to be an overkill. Since we are not interested in the exact realization of the entire history $S_1, S_2, \ldots, S_n$ (i.e., S_n only captures the number of flips required to get a total of n heads), we can generate a random variable with the same distribution as S_n with far fewer flips. In fact, the expected number of flips required can be lower-bounded by $H(S_n)$ and upper-bounded by $H(S_n) + 2$ (Theorem 5.12.3, page 115). S_n has a negative binomial distribution; i.e., $Pr(S_n = k) = \left(\begin{array}{c} k-1 \\ n-1 \end{array} \right) (\frac{1}{2})^k$. Computing $H(S_n)$ is difficult (and fruitless in the exam setting). Note, however, that for large n, the distribution of S_n will tend to Gaussian with mean $\frac{n}{p}$ and variance $np(1-p)$. You were not required to evaluate $H(S_n)$ to receive full credit, though.

In short, to receive full credit for this problem, you simply need to show that the expected number of flips required is between $H(S_n)$ an $H(S_n) + 2$, and set up the expression of $H(S_n)$ in terms of the pmf of S_n.

Here is a specific example for $n = 100$. Based on earlier discussion, $Pr(S_{100} = k) = \left(\begin{array}{c} k-1 \\ 100-1 \end{array} \right) (\frac{1}{2})^k$. The corresponding $H(S_n)$ is 5.8636. Thus, the expected number of flips required is somewhere between 5.8636 and 7.8636, substantially fewer than the popular answer of $2 * n = 2 * 100 = 200$.

8. **Entropy rate.**
Let $\{X_i\}$ be a stationary stochastic process with entropy rate $H(\mathcal{X})$.

(a) Argue that $H(\mathcal{X}) \le H(X_1)$.

(b) What are the conditions for equality?

Solution: Entropy Rate.

(a) From Theorem 4.2.1

$$H(\mathcal{X}) = H(X_1|X_0, X_{-1}, \ldots) \leq H(X_1) \tag{3}$$

since conditioning reduces entropy.

(b) We have equality only if X_1 is independent of the past $X_0, X_{-1}, \ldots$. Since it is stated that (X_i) is a stationary process, the X_i are identically distributed as well. Thus, we have equality if and only if X_i is an i.i.d. process.