Homework Set #2
Markov Process, Gambling, Causal conditioning

1. **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

$$\begin{aligned}\frac{1}{2n} I(X_1^n; X_{n+1}^{2n}) &= \frac{1}{2n}\big(H(X_1^n) + H(X_{n+1}^{2n}) - H(X_1^{2n})\big)\\ &= \frac{1}{n}H(X_1^n) - \frac{1}{2n}H(X_1^{2n}) \to 0.\end{aligned}$$

2. **Functions of a stochastic process.**

(a) Consider a stationary stochastic process $X_1, X_2, \ldots, X_n$, and let $Y_1, Y_2, \ldots, Y_n$ be defined by

$$Y_i = f(X_i), \qquad i = 1, 2, \ldots,$$

for some function f. Prove that

$$H(\mathcal{Y}) \le H(\mathcal{X}).$$

(b) What is the relationship between the entropy rate $H(\mathcal{Z})$ and $H(\mathcal{X})$ if

$$\begin{aligned}Z_1 &= g_1(X_1),\\ Z_i &= g_2(X_i, X_{i-1}), \qquad i = 2, 3, \ldots,\end{aligned}$$

for some functions g_1 and g_2.

Solution: Functions of a stochastic process.

(a) Since Y^n is a function of X^n for all n, we have $H(Y^n) \leq H(X^n)$ for all n. Thus,

$$\lim_n \frac{1}{n} H(Y^n) \leq \lim_n \frac{1}{n} H(X^n).$$

(The limit on the left exists since Y_i, $i = 1, 2, \ldots$ is stationary.)

(b) For the same reason, we have

$$\lim_n \frac{1}{n} H(Z^n) \leq \lim_n \frac{1}{n} H(X^n).$$

To justify the limit on the left is well-defined, we observe that $Z_2, Z_3, \ldots$ is stationary, and that $H(Z^n) = H(Z_2^n) + H(Z_1|Z_2^n)$, so that

$$\frac{1}{n} H(Z^n) \to \lim_n \frac{1}{n} H(Z_2^n) + \lim_n \frac{1}{n} H(Z_1|Z_2^n) = \lim_n \frac{1}{n} H(Z_2^n).$$

3. **Entropy rates of Markov chains.**

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}.$$

(b) What values of p_{01}, p_{10} maximize the entropy rate?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 1 & 0 \end{bmatrix}.$$

(d) Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of p should be less than $\frac{1}{2}$, since the 0 state permits more information to be generated than the 1 state.

Solution: Entropy rates of Markov chains.

(a) The stationary distribution is easily calculated. (See EIT pp. 62–63.)

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

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

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

(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.

4. **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,

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

(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,

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

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

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

(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.

5. **Horse race.** Three horses run a race. A gambler offers 3-for-1 odds on each of the horses. These are fair odds under the assumption that all horses are equally likely to win the race. The true win probabilities are known to be

$$\mathbf{p} = (p_1, p_2, p_3) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right). \tag{1}$$

Let $\mathbf{b} = (b_1, b_2, b_3)$, $b_i \geq 0$, $\sum b_i = 1$, be the amount invested on each of the horses. The expected log wealth is thus

$$W(\mathbf{b}) = \sum_{i=1}^{3} p_i \log 3b_i. \tag{2}$$

(a) Maximize this over $\mathbf{b}$ to find $\mathbf{b}^*$ and W^*. Thus the wealth achieved in repeated horse races should grow to infinity like 2^{nW^*} with probability one.

(b) Show that if instead we put all of our money on horse 1, the most likely winner, we will eventually go broke with probability one.

Solution: Horse race.

(a) The doubling rate

$$\begin{aligned}
W(\mathbf{b}) &= \sum_i p_i \log b_i o_i \\
&= \sum_i p_i \log 3b_i \\
&= \sum p_i \log 3 + \sum p_i \log p_i - \sum p_i \log \frac{p_i}{b_i} \\
&= \log 3 - H(\mathbf{p}) - D(\mathbf{p}||\mathbf{b}) \\
&\leq \log 3 - H(\mathbf{p}),
\end{aligned}$$

with equality iff $\mathbf{p} = \mathbf{b}$. Hence $\mathbf{b}^* = \mathbf{p} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $W^* = \log 3 - H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = \frac{1}{2} \log \frac{9}{8} = 0.085$.
By the strong law of large numbers,

$$\begin{aligned}
S_n &= \prod_j 3b(X_j) \\
&= 2^{n(\frac{1}{n}\sum_j \log 3b(X_j))} \\
&\to 2^{nE \log 3b(X)} \\
&= 2^{nW(\mathbf{b})}
\end{aligned}$$

When $\mathbf{b} = \mathbf{b}^*$, $W(\mathbf{b}) = W^*$ and $S_n \doteq 2^{nW^*} = 2^{0.085n} = (1.06)^n$.

(b) If we put all the money on the first horse, then the probability that we do not go broke in n races is $(\frac{1}{2})^n$. Since this probability goes to zero with n, the probability of the set of outcomes where we do not ever go broke is zero, and we will go broke with probability 1.
Alternatively, if $\mathbf{b} = (1, 0, 0)$, then $W(\mathbf{b}) = -\infty$ and

$$S_n \to 2^{nW} = 0 \quad \text{w.p.1} \tag{3}$$

by the strong law of large numbers.

6. **Memoryless channel without feedback.** Show that for a memoryless channel $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ without feedback

$$P(y^n|x^n) = P(y^n||x^n), \quad \forall\, n, x^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n. \tag{4}$$

Hint: recall the definition of *memoryless* channel:

$$P(y_i|x^i, y^{i-1}) = P(y_i|x_i), \quad \forall\, i, x^i \in \mathcal{X}^i,\ y^i \in \mathcal{Y}^i, \tag{5}$$

or equivalently,

$$P(y^n||x^n) = \prod_{i=1}^{n} P(y_i|x_i), \quad \forall\, n, x^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n. \tag{6}$$

Recall also that communication without feedback implies

$$P(x^n||y^{n-1}) = P(x^n), \quad \forall\, n, x^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n. \tag{7}$$

Solution: Memoryless channel without feedback.

$$\begin{aligned}
P(y^n|x^n) &= \prod_{i=1}^{n} P(y_i|x^n, y^{i-1}) \\
&= \prod_{i=1}^{n} P(y_i|x^i, x_{i+1}^n, y^{i-1}) \\
&= \prod_{i=1}^{n} \frac{P(x_{i+1}^n|x^i, y^i)P(y_i|x^i, y^{i-1})P(x^i, y^{i-1})}{P(x_{i+1}^n|x^i, y^{i-1})P(x^i, y^{i-1})} \\
&= \prod_{i=1}^{n} \frac{P(x_{i+1}^n|x^i)P(y_i|x^i, y^{i-1})}{P(x_{i+1}^n|x^i)} \\
&= \prod_{i=1}^{n} P(y_i|x^i, y^{i-1}) \\
&= \prod_{i=1}^{n} P(y_i|x_i) \\
&= P(y^n||x^n)
\end{aligned}$$

where $P(x_{i+1}^n|x^i, y^i) = P(x_{i+1}^n|x^i, y^{i-1}) = P(x_{i+1}^n|x^i)$ since there is no feedback.

7. **Alternative presentation of Directed information** Show the following identity

$$I(Y^n \to X^n) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i; X_i^n|X^{i-1}, Y^{i-1}) \tag{8}$$

Solution: Alternative presentation of Directed information.

$$I(Y^n \to X^n) \triangleq H(X^n) - H(X^n||Y^n)$$

$$\begin{aligned}
\sum_{i=1}^{n} I(Y_i; X_i^n|X^{i-1}, Y^{i-1}) &= \sum_{i=1}^{n} \left[H(Y_i|Y^{i-1}.X^{i-1}) - H(Y_i|Y^{i-1}, X^n)\right] \\
&= \sum_{i=1}^{n} \left[H(Y_i; X_i|Y^{i-1}, X^{i-1}) - H(X_i|Y^i, X^{i-1})\right] - H(Y^n|X^n) \\
&= H(Y^n; X^n) - \sum_{i=1}^{n} H(X_i|Y^i, X^{i-1}) - H(Y^n|X^n) \\
&= H(X^n) - \sum_{i=1}^{n} H(X_i|Y^i, X^{i-1}) \\
&= H(X^n) - H(X^n||Y^n) \\
&= I(Y^n \to X^n)
\end{aligned}$$

8. **Betting in a Markov horse race process with causal side information.** Consider the case in which two horses are racing, and the winning horse, X_i behaves as a Markov process as shown in Figure 1. A horse that won will win again with probability p and lose with probability $1-p$. At time zero, we assume that both horses have probability $1/2$ of wining. The side information revealed to the gambler at time i is Y_i, which is a noisy observation of the horse race outcome X_i. It has probability $1-q$ of being equal to X_i, and q of being different from X_i. In other words, $Y_i = X_i + V_i \mod 2$, where V_i is an i.i.d process, $\sim Bernoulli(q)$.

   Show that the increase in growth rate due to side information $\Delta W := \frac{1}{n}\Delta W(X^n||Y^n)$ is
   $$\Delta W = h(p * q) - h(q), \tag{9}$$
   where the function $h(\cdot)$ denotes the binary entropy, i.e., $h(x) = -x\log x - (1-x)\log(1-x)$, and $p * q$ denotes the parameter of a Bernoulli distribution that results from convolving two Bernoulli distributions with parameters p and q, i.e., $p * q = (1-p)q + (1-q)p$.

   Hint: you may use the identity from previous question $I(Y^n \to X^n) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i; X_i^n|X^{i-1}, Y^{i-1})$.

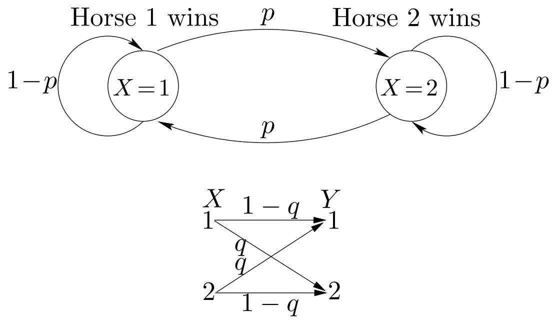

Figure 1: The winning horse X_i is represented as a Markov process with two states. In state 1, horse number 1 wins, and in state 2, horse number 2 wins. The side information, Y_i, is a noisy observation of the winning horse, X_i.

Solution: Betting in a Markov horse race process with causal side information.

Due to stationarity

$$
\begin{aligned}
\Delta W &= \lim_{n\to\infty} \frac{1}{n} I(Y^n \to X^n) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y^i|X^{i-1}) - H(Y^i|X^i) \right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y^i|X_{i-1}) - H(Y_2^i|X_2^i) - H(Y_1|X_1) \right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y^i|X_{i-1}) - H(Y^{i-1}|X^{i-1}) - H(Y_1|X_1) \right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y^{i-1}, X_{i-1}) - H(Y_1|X_1) \right] \\
&= H(Y_1|X_0) - H(Y_1|X_1) \\
&= h(p * q) - h(q)
\end{aligned}
$$