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.

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

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

(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.

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.

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 $n \geq 2$?

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.

6. **Alternative representation of directed information** Show the following identity

$$I(Y^n \to X^n) \quad = \quad \sum_{i=1}^{n} I(Y_i; X_i^n | X^{i-1}, Y^{i-1}) \tag{3}$$

7. **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, $\sim \mathit{Bernoulli}(q)$.

   Show that the increase in growth rate due to side information $\Delta W := \frac{1}{n}\Delta W(X^n||Y^n)$ is

$$\Delta W = h(p * q) - h(q), \tag{4}$$

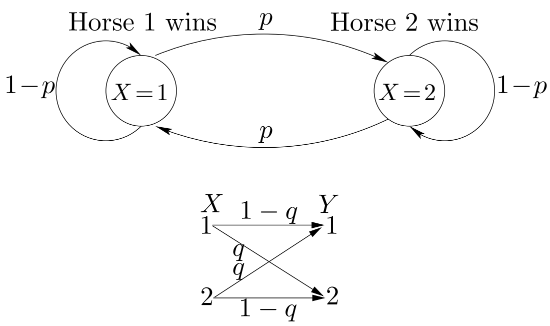

Figure 1: The winning horse X_i is represented as a Markov process with two states. In state 1, horse number 1 wins, and in state 2, horse number 2 wins. The side information, Y_i, is a noisy observation of the winning horse, X_i.

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

8. **Channel without feedback.** Show that for a channel $P_{Y^n||X^n}$ without feedback

$$P(y^n|x^n) = P(y^n||x^n), \tag{5}$$

and in particular if the channel is memoryless

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x_i), \tag{6}$$

for all $n, x^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n$.

Hint: recall the definition of *memoryless* channel:

$$P(y_i|x^i, y^{i-1}) = P(y_i|x_i), \tag{7}$$

or equivalently,

$$P(y^n||x^n) = \prod_{i=1}^{n} P(y_i|x_i) \tag{8}$$

Recall also that communication without feedback implies

$$P(x^n||y^{n-1}) = P(x^n) \tag{9}$$