Problem 1 (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 \( 1/2 \), since the 0 state permits more information to be generated than the 1 state.
(e) Let \( N(t) \) be the number of allowable state sequences of length \( t \) for the Markov chain of part (c). Find \( N(t) \) and calculate
\[ H_0 = \lim_{t \to \infty} \frac{1}{t} \log N(t). \]
Hint: Find a linear recurrence that expresses \( N(t) \) in terms of \( N(t-1) \) and \( N(t-2) \). Why is \( H_0 \) an upper bound on the entropy rate of the Markov chain? Compare \( H_0 \) with the maximum entropy found in part (d).

Solution.

(a) The stationary distribution is easily calculated.
\[ \mu_0 = \frac{p_{10}}{p_{01} + p_{10}}, \quad \mu_1 = \frac{p_{01}}{p_{01} + p_{10}} \]
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 \( P(X_i = 0) = Pr(X_i = 1) = 1/2 \).

(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
\[
H(p) = H(1 - p) = H\left(\frac{\sqrt{5} - 1}{2}\right) = 0.694.
\]

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. Consider any allowable sequence of symbols of length \( t \). If the first symbol is 1, then the next symbol must be 0; the remaining \( N(t-2) \) symbols can form any allowable sequence. If the first symbol is 0, then the remaining \( N(t-1) \) symbols can be any allowable sequence. So the number of allowable sequences of length \( t \) satisfies the recurrence
\[
N(t) = N(t-1) + N(t-2), \quad N(1) = 2, N(2) = 3
\]
(The initial conditions are obtained by observing that for \( t = 2 \) only the sequence 11 is not allowed. We could also choose \( N(0) = 1 \) as an initial condition, since there is exactly one allowable sequence of length 0, namely, the empty sequence.) The sequence \( N(t) \) grows exponentially, that is, \( N(t) \approx c\lambda^t \). In fact
\[
N(t) = \left(\frac{-3\sqrt{5} + 1}{10} + \frac{1}{2}\right) \left(\frac{-\sqrt{5} + 1}{2}\right)^t + \left(\frac{1}{2} + 3\sqrt{5}\right) \left(\frac{\sqrt{5} + 1}{2}\right)^t.
\]
\[
H_0 = \lim_{t \to \infty} \frac{1}{t} \log N(t) = \lim_{t \to \infty} \frac{1}{t} \log_2 N(t) = \log_2 \frac{1 + \sqrt{5}}{2}.
\]

Problem 2 (Maximum entropy process) A discrete memoryless source has alphabet \( \{1, 2\} \) where the symbol 1 has duration 1 and the symbol 2 has duration 2. The probabilities of 1 and 2 are \( p_1 \) and \( p_2 \), respectively. Find the value of \( p_1 \) that maximizes the source entropy per unit time \( H(X)/E(l(X)) \). What is the maximum value \( H \)?
Solution. The entropy per symbol of the source is
\[
H(p_1) = -p_1 \log p_1 - (1 - p_1) \log(1 - p_1).
\]
and the average symbol duration (or time per symbol) is
\[
T(p_1) = 1p_1 + 2p_2 = p_1 + 2(1 - p_1) = 2 - p_1 = 1 + p_2.
\]
Therefore the source entropy per unit time is
\[
f(p_1) = \frac{H(p_1)}{T(p_1)} = \frac{-p_1 \log p_1 - (1 - p_1) \log(1 - p_1)}{2 - p_1}
\]
Since \(f(0) = f(1) = 0\), the maximum value of \(f(p_1)\) must occur for some point \(p_1\) such that \(0 < p_1 < 1\) and \(\frac{df}{dp_1} = 0\).

\[
\frac{d}{dp_1} \left( \frac{-p_1 \log p_1 - (1 - p_1) \log(1 - p_1)}{2 - p_1} \right) = \frac{1}{(p_1 - 2)^2} \left( \ln (1 - p_1) - 2 \ln p_1 \right) = 0
\]

\[1 - p_1 = \frac{p_1^2}{2}, \quad \text{Solution is: } \frac{1}{2} \sqrt{5} - \frac{1}{2}, -\frac{1}{2} \sqrt{5} - \frac{1}{2}. \text{The corresponding entropy per unit time is}
\]
\[
f(p_1) = \frac{-\left(1 + \frac{p_1^2}{2}\right) \log p_1}{1 + \frac{p_1^2}{2}} = -\log p_1 = 0.69424 \text{ bits.}
\]

Problem 3 (Initial conditions) Show, for a Markov chain, that
\[
H(X_0 | X_n) \geq H(X_0 | X_{n-1}).
\]
Thus initial conditions \(X_0\) become more difficult to recover as the future \(X_n\) unfolds.

Solution. For a Markov chain, by the data processing theorem, we have
\[
I(X_0; X_{n-1}) \geq I(X_0; X_n).
\]
Therefore
\[
H(X_0) - H(X_0 | X_{n-1}) \geq H(X_0) - H(X_0 | X_n),
\]
i.e. \(H(X_0 | X_n)\) increases with \(n\).