Solution of Problem 1

\( \mathbf{p} = (p_1, p_2) \) be the stationary distribution for a two state homogeneous Markov chain with states \( \{0, 1\} \) and transition matrix \( \Pi = \begin{pmatrix} 1 - \alpha & 1 - \beta \\ \beta & \alpha \end{pmatrix} \).

We know, for a stationary distribution \( \mathbf{p} \Pi = \mathbf{p} \). We also know \( p_1 + p_2 = 1 \) i.e., \( p_1 = 1 - p_2 \).

\[
\begin{align*}
(p_1, p_2) \Pi &= (p_1, p_2) \\
(p_1, p_2) \begin{pmatrix} 1 - \alpha & 1 - \beta \\ \beta & \alpha \end{pmatrix} &= (p_1, p_2)
\end{align*}
\]

(1)

We get

\[
\begin{align*}
(1 - \alpha)p_1 + \beta p_2 &= p_1 \\
(1 - \beta)p_1 + \alpha p_2 &= p_2
\end{align*}
\]

(2)

substituting \( p_1 = 1 - p_2 \) in one of the above equations, we get

\[
\begin{align*}
(1 - \alpha)p_1 + \beta(1 - p_1) &= p_1 \\
p_1 &= \frac{\beta}{\alpha + \beta}
\end{align*}
\]

(3)

and

\[
\begin{align*}
p_2 &= 1 - p_1 \\
p_2 &= \frac{\alpha}{\alpha + \beta}
\end{align*}
\]

(4)

Hence the stationary distribution \( \mathbf{p} = (\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta}) \).

Solution of Problem 2

Let \( X_0, X_1, X_2, \ldots X_n \) are drawn i.i.d \( \sim p(x), x \in X = \{1, 2, 3, \ldots, m\} \), and the waiting time to the next occurrence of \( X_0 \) has a geometric distribution with probability of success \( p(x_0) \).

\( a) \) Given \( X_0 = i \). \( P(X_n = i) = (1 - p(i))^{n-1} p(i) \).
\[ E[N|X_0 = i] = \sum_{n=1}^{\infty} n(1-p(i))^{n-1}p(i) \]
\[ = \sum_{\tilde{n}=0}^{\infty} (\tilde{n}+1)(1-p(i))^{\tilde{n}}p(i) \quad \text{(when } \tilde{n} = n-1) \quad (5) \]
\[ = p(i)\sum_{\tilde{n}=0}^{\infty} (\tilde{n})(1-p(i))^{\tilde{n}} + p(i)\sum_{\tilde{n}=0}^{\infty} (1-p(i))^{\tilde{n}} \]
Using the given hint, For \(0 < r < 1\) we have
\[ \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \quad \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}. \]
we can write
\[ E[N|X_0 = i] = p(i)\frac{(1-p(i))}{(p(i))^2} + p(i)\frac{1}{p(i)} \quad (6) \]
\[ = \frac{(1-p(i))}{p(i)} + 1 = \frac{1}{p(i)}. \]
Therefore,
\[ EN = E[E[N|X_0 = i]] = \sum_{i=1}^{m} P(X_0 = i)E[N|X_0 = i] = \sum_{i=1}^{m} p(i)\frac{1}{p(i)} = m. \quad (7) \]

b) From (a), we know, \(E[N|X_0 = i] = \frac{1}{p(i)}\).
\[ E \log N = \sum_{i=1}^{m} P(X_0 = i)E[\log N|X_0 = i] \]
\[ \leq \sum_{i=1}^{m} P(X_0 = i) \log E[N|X_0 = i] \quad (\text{Jensen’s Inequality}) \quad (8) \]
\[ = \sum_{i=1}^{m} p(i) \log \frac{1}{p(i)} \]
\[ = H(X). \]
Hence, we get \(E \log N \leq H(X)\).

**Solution of Problem 3**

a) By the chain rule, we can write
\[ H(X_1, X_2, \ldots, X_n) = \sum_{i=0}^{n} H(X_i|X_{i-1}, \ldots, X_0) \]
\[ = H(X_0) + H(X_1|X_0) + \sum_{i=2}^{n} H(X_i|X_{i-1}, X_{i-2}) \quad (9) \]
Since for \(i > 1\), the next position depends only on the previous two i.e., the dog’s walk is 2nd order Markov, if the dog’s position is the state.
Since \(X_0 = 0\) deterministically, \(H(X_0) = 0\).
For the first step, it is equally likely to be positive or negative, \( H(X_1|X_0) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1 \).

Furthermore, for \( i > 1 \),

\[
H(X_i|X_{i-1}, X_{i-2}) = H(0.1, 0.9). \tag{10}
\]

So,

\[
H(X_1, X_2, ..., X_n) = 1 + (n - 1)H(0.1, 0.9). \tag{11}
\]

b) The entropy rate of the dog:

\[
\frac{1}{n+1} H(X_0, X_1, ..., X_n) = \frac{1 + (n - 1)H(0.1, 0.9)}{n+1} \rightarrow_{n \to \infty} H(0.1, 0.9) \tag{12}
\]

c) The dog must take at least one step to establish the direction of travel from which it ultimately reverses. Letting \( S \) be the number of steps taken between reversals, we have

\[
E(S) = \sum_{s=1}^{\infty} s(0.9)^{s-1}(0.1) \tag{13}
\]

Starting at time 0, the expected number of steps to the first reversal is 11.

**Solution of Problem 4**

Given:

- \( X_i \) be i.i.d \( \sim p(x) \), \( x \in \mathcal{X} = \{1, 2, 3, ..., m\} \).
- \( \mu = EX \) and \( H = -\sum p(x) \log p(x) \).
- The typical set \( A^n_\epsilon = \{(x_1, x_2, ..., x_n) \in \mathcal{X}^n : | -\frac{1}{n} \log p(x_1, x_2, ..., x_n) - H| \leq \epsilon\} \).
- \( B^n_\epsilon = \{(x_1, x_2, ..., x_n) \in \mathcal{X}^n : \left| \frac{1}{n} \sum_{i=1}^{n} x_i - \mu \right| \leq \epsilon\} \).

a) Yes, By the definition of AEP for discrete random variables, the probability \((X_1, X_2, ..., X_n)\) belongs to a typical set goes to 1 as \( n \to \infty \).

b) Yes, by the strong law of large numbers \( P((X_1, X_2, ..., X_n) \in B^n_\epsilon) \to 1 \).

For any \( \epsilon > 0 \), there exists \( N_1 \) such that \( P((X_1, X_2, ..., X_n) \in A^n_\epsilon) > 1 - \frac{\epsilon}{2} \) for all \( n > N_1 \).

Similarly, we can say that there exists \( N_2 \) such that \( P((X_1, X_2, ..., X_n) \in B^n_\epsilon) > 1 - \frac{\epsilon}{2} \) for all \( n > N_2 \).

So for all \( n > \max(N_1, N_2) \):

\[
P((X_1, X_2, ..., X_n) \in A^n_\epsilon \cap B^n_\epsilon) = P((X_1, X_2, ..., X_n) \in A^n_\epsilon) + P((X_1, X_2, ..., X_n) \in B^n_\epsilon) - P((X_1, X_2, ..., X_n) \in A^n_\epsilon \cup B^n_\epsilon) \]

\[
> 1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 \]

\[
= 1 - \epsilon. \tag{14}
\]

So for any \( \epsilon > 0 \), there exists \( N = \max(N_1, N_2) \) such that \( P((X_1, X_2, ..., X_n) \in A^n_\epsilon \cap B^n_\epsilon) > 1 - \epsilon \) for all \( n > N \), therefore \( P((X_1, X_2, ..., X_n) \in A^n_\epsilon \cap B^n_\epsilon) \to 1 \).
c) By the law of total probability, we get\
\[ \sum_{(x_1, x_2, \ldots, x_n) \in A^n \cap B^n} p(x_1, x_2, \ldots, x_n) \leq 1. \]
For \((x_1, x_2, \ldots, x_n) \in A^n\), from Theorem 2.4.4, we get \(p(x_1, x_2, \ldots, x_n) \geq 2^{-n(H+\epsilon)}\).
Using these two equations, we can write
\[
1 \geq \sum_{(x_1, x_2, \ldots, x_n) \in A^n \cap B^n} p(x_1, x_2, \ldots, x_n) \geq \sum_{(x_1, x_2, \ldots, x_n) \in A^n \cap B^n} 2^{-n(H+\epsilon)} = |A^n \cap B^n| 2^{-n(H+\epsilon)}.
\] (15)
Multiplying through \(2^{n(H+\epsilon)}\), we get \(|A^n \cap B^n| \leq 2^{n(H+\epsilon)}\).

\[
\frac{1}{2} \leq \sum_{(x_1, x_2, \ldots, x_n) \in A^n \cap B^n} p(x_1, x_2, \ldots, x_n) \leq \sum_{(x_1, x_2, \ldots, x_n) \in A^n \cap B^n} 2^{-n(H-\epsilon)} = |A^n \cap B^n| 2^{-n(H-\epsilon)}.
\] (16)
Multiplying through \(2^{n(H-\epsilon)}\), we get \(|A^n \cap B^n| \geq (\frac{1}{2})2^{n(H-\epsilon)}\) for sufficiently large \(n\).

**Solution of Problem 5**

\[
\frac{1}{n} \log \frac{p(X_1, X_2, \ldots, X_n)p(Y_1, Y_2, \ldots, Y_n)}{p(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n)} = \frac{1}{n} \log \prod_{i=1}^{n} \frac{p(X_i)p(Y_i)}{p(X_i, Y_i)}
= \frac{1}{n} \sum_{i=1}^{n} \log \frac{p(X_i)p(Y_i)}{p(X_i, Y_i)}
\xrightarrow{n \to \infty} E \log \frac{p(X_i)p(Y_i)}{p(X_i, Y_i)}
= -I(X; Y)
\] (17)
Hence, we get \(p(X_1, X_2, \ldots, X_n)p(Y_1, Y_2, \ldots, Y_n) = 2^{-nI(X; Y)}\), which will converge to 1 if \(X\) and \(Y\) are independent.