Problem 3.1:

(a) Let X have cummulative distribution function $F(x)$, then

$$\begin{aligned} E[X] &= \int_0^\infty x\, dF(x) \\ &= \underbrace{\int_0^\delta x\, dF(x)}_{\geq 0} + \int_\delta^\infty x\, dF(x) \\ &\geq \int_\delta^\infty x\, dF(x) \\ &\geq \int_\delta^\infty \delta\, dF(x) \\ &= \delta\, Pr\{X \geq \delta\}. \end{aligned}$$

Thus

$$Pr\{X \geq \delta\} \leq \frac{E[X]}{\delta}.$$

Given $\delta > 0$, an example of a R.V. with

$$Pr\{X \geq \delta\} = \frac{E[X]}{\delta}$$

is:

$$X = \begin{cases} \delta & \text{with probability } \frac{\mu}{\delta}, \\ 0 & \text{with probability } 1 - \frac{\mu}{\delta}, \end{cases}$$

where $\mu \triangleq E[X]$ and $\frac{\mu}{\delta} \leq 1$ (i.e., $\mu \leq \delta$). □

(b) Letting $X = (Y-\mu)^2$ in Markov's inequality (with $\delta = \varepsilon^2$) yields:

$$\begin{aligned} Pr\{(Y-\mu)^2 > \varepsilon^2\} &\leq Pr\{(Y-\mu)^2 \geq \varepsilon^2\} \\ &\leq \frac{E[(Y-\mu)^2]}{\varepsilon^2} \\ &= \frac{\sigma^2}{\varepsilon^2}. \end{aligned}$$

But, $Pr\{(Y-\mu)^2 > \varepsilon^2\} = Pr\{|Y-\mu| > \varepsilon\}$. Thus, Chebyshev's inequality is proved. □

(c) Use Chebyshev's inequality with $Y = \overline{Z}_n$ along with the facts that $E[\overline{Z}_n] = \mu$ and $Var(\overline{Z}_n) = \frac{\sigma^2}{n}$, (since $\overline{Z}$ is the sum of n iid R.V.'s $\frac{Z_i}{n}$, each with variance $\frac{\sigma^2}{n^2}$). Hence,

$$Pr\{|\overline{Z}_n - \mu| > \varepsilon\} \leq \frac{Var(\overline{Z}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Taking $n \to \infty$, yields the Weak Law of Large Numbers (WLLN). □

Problem 3.3: Let $\mathcal{X}$ be a DMS with $P(X=0) = 0.995$ and $\mathcal{X} = \{0, 1\}$.

(a) The number of sequences of length 100 with 3 or fewer ones is:

$$\begin{aligned}\sum_{i=0}^{3}\binom{100}{i} &= \binom{100}{0}+\binom{100}{1}+\binom{100}{2}+\binom{100}{3}\\ &= 1+100+4950+161700\\ &= 166751.\end{aligned}$$

Thus, the codeword length $= \lceil \log_2 166751 \rceil = \lceil 17.34 \rceil = 18$.

So the block encoder is: $f : \{0,1\}^{100} \longrightarrow \{0,1\}^{18}$, and is uniquely decodable on the 100-bit sequence with 3 or fewer ones.

Note: Code rate $R = \frac{18}{100} = 0.18$ bits/source sample, while the source entropy $H(X) = h_b(0.995) = 0.0454$ bits/source sample. $R >> H(X)$: Not a very good code.

(b) Let $C^{(100)}$ be the set of source sequences that have no codewords assigned to them:

$$\begin{aligned}Pr\{C^{(100)}\} &= 1 - Pr\{\text{100-bit sequence has 3 or fewer ones}\}\\ &= 1-\sum_{i=0}^{3}\binom{100}{i}(0.005)^i(0.995)^{100-i}\\ &= 1-0.99833 = 0.00167.\end{aligned}$$

(c) Let $S_n = X_1 + \ldots + X_n$ and let μ and σ^2 be the mean and variance of each of the iid R.V.'s X_i's. $\mu = 0.005$ and $\sigma^2 = 0.004975$. Here $n = 100$. Note that $E[S_n] = n\mu$. By Chebyshev:

$$Pr\{|S_n - n\mu| \geq \varepsilon\} \leq \frac{Var[S_n]}{\varepsilon^2} = \frac{n\sigma^2}{\varepsilon^2}.$$

Need to find $Pr\{S_{100} \geq 4\}$!

But $S_{100} \geq 4 \Leftrightarrow |S_{100} - (100)(0.005)| \geq 3.5$. Thus choose $\varepsilon = 3.5$, and

$$Pr\{C^{(100)}\} = Pr\{S_{100} \geq 4\} \leq \frac{(100)(0.004975)}{(3.5)^2} = 0.04061.$$

Compared to the actual probability from part (b), this is a pretty loose bound !

Problem 3.5: Let $X_1, X_2, \ldots$ be a sequence of iid RV's with generic pmf $p(x)$, where $x \in \mathcal{X} = \{1, 2, \ldots, m\}$. Let $q(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} q(X_i)$, where $q(\cdot)$ is another pmf on $\mathcal{X}$.

(a) We first observe that

$$-\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) = -\frac{1}{n}\log\left[\prod_{i=1}^{n} q(X_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\left[-\log q(X_i)\right].$$

Since the X_i's are iid, then so are the RV's $Y_i \triangleq [-\log q(X_i)]$, for $i = 1, 2, \ldots$. Therefore, by the WLLN, we have that $\frac{1}{n}\sum_{i=1}^{n} Y_i$ converges (in probability) to $E[Y_1] = E_p[-\log q(X_1)]$ as $n \to \infty$; (in fact, by the strong law of large numbers (SLLN), the convergence is with probability one):

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) &= E_p[-\log q(X_1)] \quad \text{with prob. 1}\\ &= \sum_{x\in\mathcal{X}} p(x)\log\frac{1}{q(x)}\\ &= \sum_{x\in\mathcal{X}} p(x)\log\left[\frac{1}{p(x)}\frac{p(x)}{q(x)}\right]\\ &= H(p) + D(p||q).\end{aligned}$$

(b) Similarly, by the SLLN, we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log\frac{q(X_1,\ldots,X_n)}{p(X_1,\ldots,X_n)} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{q(X_i)}{p(X_i)}\right] \\
&= E_p\left[\log\frac{q(X_1)}{p(X_1)}\right] \quad \text{with prob. 1} \\
&= \sum_{x\in\mathcal{X}} p(x)\log\frac{q(x)}{p(x)} \\
&= -D(p||q).
\end{aligned}$$

Problem 4.4:

(a) By the chain rule for entropy

$$\frac{1}{n}H(X_1,\ldots,X_n) = \frac{1}{n}[H(X_1,\ldots,X_{n-1}) + H(X_n|X_1,\ldots,X_{n-1})] \quad ^{(*)},$$

but by stationarity and the fact that conditioning decreases entropy,

$$H(X_n|X_1,\ldots,X_{n-1}) \le H(X_i|X_1,\ldots,X_{i-1}) \ \ \forall\, 1\le i\le n$$

Thus

$$\begin{aligned}
(n-1)H(X_n|X_1,\ldots,X_{n-1}) &= \sum_{i=1}^{n-1} H(X_n|X_1,\ldots,X_{n-1}) \\
&\le \sum_{i=1}^{n-1} H(X_i|X_1,\ldots,X_{i-1}) \\
&= H(X_1,\ldots,X_{n-1}) \quad \text{by the chain rule.}
\end{aligned}$$

Hence

$$H(X_n|X_1,\ldots,X_{n-1}) \le \frac{1}{n-1}H(X_1,\ldots,X_{n-1}).$$

Now applying the above inequality to (*) yields

$$\begin{aligned}
\frac{1}{n}H(X_1,\ldots,X_n) &\le \frac{1}{n}[H(X_1,\ldots,X_{n-1}) + \frac{1}{n-1}H(X_1,\ldots,X_{n-1})] \\
&= \left(\frac{1}{n}\right)\left(\frac{n}{n-1}\right)H(X_1,\ldots,X_{n-1}).
\end{aligned}$$

Thus

$$\frac{1}{n}H(X_1,\ldots,X_n) \le \frac{1}{n-1}H(X_1,\ldots,X_{n-1}).$$

□

(b) As in (a) we know that

$$H(X_n|X_1,\ldots,X_{n-1}) \le H(X_i|X_1,\ldots,X_{i-1}) \ \ \forall\, 1\le i\le n.$$

Write

$$\begin{aligned}
H(X_n|X_1,\ldots,X_{n-1}) &= \frac{1}{n}\sum_{i=1}^{n} H(X_n|X_1,\ldots,X_{n-1}) \\
&\stackrel{(1)}{\le} \frac{1}{n}\sum_{i=1}^{n} H(X_i|X_1,\ldots,X_{i-1}) \\
&\stackrel{(2)}{=} \frac{1}{n}H(X_1,\ldots,X_n),
\end{aligned}$$

where (1) follows from the above fact and (2) follows from the chain rule for entropy. □

Problem 4.5: We first determine the stationary distribution: $p_\infty(0) = \Pr\{X_n = 0\}$ and $p_\infty(1) = 1-p_\infty(0) = \Pr\{X_n = 1\}$. To do this, we solve the equation: $p_\infty(0) = (1-p_{01})p_\infty(0)+p_{10}(1-p_\infty(0))$. The solution of this is $p_\infty(0) = p_{10}/(p_{01}+p_{10})$.

(a) The entropy rate of the process is:

$$\begin{aligned} H(\mathcal{X}) &= H(X_2|X_1) \\ &= H(X_2|X_1=0)\Pr\{X_1=0\} + H(X_2|X_1=1)\Pr\{X_1=1\} \\ &= h_b(p_{01})p_\infty(0) + h_b(p_{10})p_\infty(1) \\ &= h_b(p_{01})\frac{p_{10}}{p_{01}+p_{10}} + h_b(p_{10})\frac{p_{01}}{p_{01}+p_{10}}. \end{aligned}$$

(b) To find the maximizing p_{01} and p_{10}, we solve the equations: $\frac{\partial H(\mathcal{X})}{\partial p_{01}} = 0$ and $\frac{\partial H(\mathcal{X})}{\partial p_{10}} = 0$. Another way (which is easier) is to realize that $H(\mathcal{X}) \leq 1$ with equality iff $\mathcal{X}$ is i.i.d. with a uniform distribution. This is possible only when $p_{01} = p_{10} = 1/2$.

(c) Here $p_{01} = p$ and $p_{10} = 1$. Hence

$$H(\mathcal{X}) = h_b(p)\frac{1}{1+p} + h_b(1)\frac{p}{1+p} = h_b(p)\frac{1}{1+p}.$$

(d) To find the maximum entropy, we solve the equation:

$$\frac{\partial H(\mathcal{X})}{\partial p} = \frac{h_b'(p)(1+p) - h_b(p)}{(1+p)^2} = 0,$$

which is equivalent to $h_b'(p)(1+p) - h_b(p) = 0$. Solving for this, we get $p^2 - 3p + 1 = 0$, yielding $p = (3-\sqrt{5})/2 = 0.382$. The maximum entropy rate is $H(\mathcal{X}) = 0.694241913$ bits.

(e) Let $\mathcal{A}_t = \{$allowable (non-zero probability) state sequences of length $t\}$. Then $N(t) = |\mathcal{A}_t|$. The first few sets in the sequence are:

$$\begin{aligned} \mathcal{A}_1 &= \{0,1\}, \\ \mathcal{A}_2 &= \{00, 01, 10\}, \\ \mathcal{A}_3 &= \{000, 001, 010, 100, 101\}, \\ \mathcal{A}_4 &= \{0000, 0001, 0010, 0100, 0101, 1000, 1001, 1010\}. \end{aligned}$$

The first few set sizes are: $N(0) = 1, N(1) = 2, N(2) = 3, N(3) = 5, N(4) = 8, N(5) = 13, N(6) = 21, \ldots$. It can be easily shown that

$$N(t) = N(t-1) + N(t-2), \qquad (\star)$$

i.e., $\{$the size of $\mathcal{A}_t\} = \{$the size of $\mathcal{A}_{t-1}\}+\{$number of terms in $\mathcal{A}_{t-1}$ which ends in $0\}$. This is known as *Fibonacci sequence.* Now, define

$$a_t = N(t)/N(t-1).$$

Plugging a_t into $(\star)$ above, we get $a_t = 1 + 1/a_{t-1}$. Finding the fixed point of this equation (by setting $a_t = a_{t-1}$), we obtain $\lim_{t\to\infty} a_t = a = (1+\sqrt{5})/2$.

Now,

$$N(t) = \left(\prod_{i=1}^{t} a_i\right) N(0).$$

Taking the log of both sides, we get

$$\frac{1}{t}\log_2 N(t) = \frac{1}{t}\sum_{i=1}^{t}\log_2 a_i + \frac{1}{t}\log_2 N(0).$$

Taking the limit as $t \to \infty$, we get

$$H_0 = \lim_{t\to\infty} \frac{1}{t} \log_2 N(t) = \log_2 \left(\frac{1+\sqrt{5}}{2} \right) = 0.694241913\text{bits}.$$

H_0 is an upper bound on the entropy rate because the terms in $\mathcal{A}_t$ are not necessarily uniformly distributed. H_0 is very close to the maximum entropy calculated in part (d), hence the upper bound is tight.

Problem 4.6: Consider a DMS with alphabet $\mathcal{X} = \{1, 2\}$, and symbol durations $l_1 = 1$ and $l_2 = 2$. Let $P\{X = 1\} \stackrel{\triangle}{=} p_1$ and $P\{X = 2\} \stackrel{\triangle}{=} p_2 = 1 - p_1$. Then

$$\frac{H(X)}{E[l_X]} = \frac{-p_1 \log_2 p_1 - (1-p_1)\log_2 (1-p_1)}{2-p_1}.$$

We need maximize $H(X)/E[l_X]$ over p_1. This is achieved by verifying that $\frac{\partial^2 (H(X)/E[l_X])}{\partial p_1^2} \leq 0$ (check it !) and then setting the first derivative to zero:

$$\frac{\partial (H(X)/E[l_X])}{\partial p_1} = 0.$$

Carrying the above computation yields

$$\begin{aligned} \log_2 (1-p_1) - 2\log_2 p_1 = 0 \quad &\Longleftrightarrow \quad \log_2 \frac{1-p_1}{p_1^2} = 0 \\ &\Longleftrightarrow \quad p_1^2 + p_1 - 1 = 0. \end{aligned}$$

The acceptable root of the above equation is $p_1^* = (-1 + \sqrt{5})/2 = 0.61803$. Thus

$$\left[\frac{H(X)}{E[l_X]} \right]_{max} = 0.69424 \quad \text{bits/unit time.}$$

Interestingly, the answer here is the same as in parts (d) and (e) of problem 4.5

Problem 4.7: Note that $X_0 \to X_{n-1} \to X_n$ form a Markov chain in that order. Therefore, by the data processing theorem, we have

$$I(X_0; X_{n-1}) \geq I(X_0; X_n);$$

i.e.,

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

Therefore $H(X_0|X_n) \geq H(X_0|X_{n-1})$ and hence, $H(X_0|X_n)$ is non-decreasing in n. Note that stationarity is NOT required here. □

Problems 4.8: First consider the model in problem 4.8. Let

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, \ldots, x_n) \in \{0,1\}^n, \\ w_H(\mathbf{x}) &= \sum_{i=1}^{n} x_i = \text{ Hamming weight of } \mathbf{x}. \end{aligned}$$

Note that $(X_1, X_2, \ldots, X_{n-1})$ are i.i.d. and $X_n = X_1 \oplus X_2 \oplus \ldots \oplus X_{n-1} = \bigoplus_{i=1}^{n-1} X_i$, where $\oplus$ is the modular-two addition. Thus

$$\Pr\{\mathbf{X} = \mathbf{x}\} = \begin{cases} (1/2)^{n-1} & \text{if } w_H(\mathbf{x}) \text{ is even} \\ 0 & \text{if } w_H(\mathbf{x}) \text{ is odd} \end{cases}.$$

(a) Note that if neither i nor j is equal to n, then X_i and X_j are independent by definition. Now assume that $i \in \{1, 2, \ldots, n-1\}$ and $j = n$.

$$\begin{aligned}
\Pr\{X_n = x_n | X_i = x_i\} &= \Pr\{\textstyle\bigoplus_{k=1}^{n-1} X_k = x_n | X_i = x_i\} \\
&= \Pr\{\textstyle\bigoplus_{k=1,k\neq i}^{n-1} X_k = x_n \oplus x_i | X_i = x_i\} \\
&= \Pr\{\textstyle\bigoplus_{k=1,k\neq i}^{n-1} X_k = x_n \oplus x_i\}.
\end{aligned}$$

The last equality is by the independence of $(X_1, X_2, \ldots, X_{n-1})$. Now, note that $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n-1})$ $\{0, 1\}^{n-2}$. Furthermore, exactly half of the sequences in $\{0, 1\}^{n-2}$ have even Hamming weight and the other half have odd Hamming weight. Thus

$$\Pr\{\textstyle\bigoplus_{k=1,k\neq i}^{n-1} X_k = 0\} = 1/2.$$

Hence, $\Pr\{X_n = x_n | X_i = x_i\} = 1/2 = \Pr\{X_n = x_n\}$, which implies that X_n and X_i are independent.

(b) $H(X_i, X_j) = H(X_i) + H(X_j) = 2$ bits if $i \neq j$.

(c)

$$H(\mathbf{X}) = H(X_1, X_2, \ldots, X_n) = \underbrace{H(X_n | X_1, \ldots, X_{n-1})}_{=0} + \underbrace{H(X_1, \ldots, X_{n-1})}_{=n-1}.$$

Thus $H(\mathbf{X}) = n - 1$ bits.

Problem 4.9: Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be a stationary process.

(a) *True*, since by the chain rule for entropy, we can write

$$\begin{aligned}
H(X_n | X_0) &= H(X_0, X_n) - H(X_0), \\
H(X_{-n} | X_0) &= H(X_{-n}, X_0) - H(X_0).
\end{aligned}$$

However $H(X_0, X_n) = H(X_{-n}, X_0)$ by stationarity; therefore $H(X_n | X_0) = H(X_{-n} | X_0)$. □

(b) *False*. Here is a counterexample. Let $X_0, X_1, \ldots, X_{n-1}$ be uniform binary iid RV's, and let $X_k = X_{k-n}$ if $k \geq n$. Then $H(X_n | X_0) = 0$ (since $X_n = X_0$), and $H(X_{n-1} | X_0) = H(X_{n-1}) = 1$. This contradicts the statement that $H(X_n | X_0) \geq H(X_{n-1} | X_0)$. □

Remark: This statement is true if the process is stationary and Markov.

(c) *True*, since

$$\begin{aligned}
H(X_n | X_1^{n-1}, X_{n+1}) &\overset{(1)}{=} H(X_1, \ldots, X_{n+1}) - H(X_1, \ldots, X_{n-1}, X_{n+1}) \\
&\overset{(2)}{=} H(X_2, \ldots, X_{n+2}) - H(X_2, \ldots, X_n, X_{n+2}) \\
&\overset{(3)}{=} H(X_{n+1} | X_2^n, X_{n+2}) \\
&\overset{(4)}{\geq} H(X_{n+1} | X_1, X_2^n, X_{n+2}) \\
&= H(X_{n+1} | X_1^n, X_{n+2}),
\end{aligned}$$

where

- (1) and (3) follow from the chain rule for entropy.
- (2) follows by stationarity.
- (4) follows from the fact that conditioning reduces entropy.

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

Problem 4.10:

(a)

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \sum_{i=3}^{n} H(X_i|X_i, \ldots, X_1).$$

For $i > 1$, $\{X_i\}$ is a *second-order* Markov chain, since at time i, we need to know only the previous two states (at times $i-2$ and $i-1$) in order to know whether the dog reversed direction or not. So

$$H(X_i|X_{i-1}, X_{i-2}, \ldots, X_1) = H(X_i|X_{i-1}, X_{i-2}).$$

Since $X_0 = 0$, X_1 takes on values $+1$ and -1 with probability $P(X_1 = +1) = P(X_1 = -1) = \frac{1}{2}$. Hence $H(X_1) = 1$.

$$\begin{aligned}
H(X_2|H_1) &= -\sum_{x_1}\sum_{x_2} p_{X_2|X_1}(x_2|x_1)p(x_1)\log p_{X_2|X_1}(x_2|x_1) \\
&= -\sum_{k=-1,+1}\ \sum_{x_2=k-1,k+1} p_{X_2|X_1}(x_2|k)p_{X_1}(k)\log p_{X_2|X_1}(x_2|k) \\
&= -\sum_{x_2=-2,0} p_{X_2|X_1}(x_2|-1)(\frac{1}{2})\log p_{X_2|X_1}(x_2|-1) \\
&\quad -\sum_{x_2=0,+2} p_{X_2|X_1}(x_2|+1)(\frac{1}{2})\log p_{X_2|X_1}(x_2|+1)
\end{aligned}$$

$$\begin{aligned}
H(X_2|H_1) &= -p_{X_2|X_1}(-2|-1)\left(\frac{1}{2}\right)\log p_{X_2|X_1}(-2|-1) \\
&\quad -p_{X_2|X_1}(0|-1)\left(\frac{1}{2}\right)\log p_{X_2|X_1}(0|-1) \\
&\quad -p_{X_2|X_1}(0|+1)\left(\frac{1}{2}\right)\log p_{X_2|X_1}(0|1) \\
&\quad -p_{X_2|X_1}(+2|+1)\left(\frac{1}{2}\right)\log p_{X_2|X_1}(+2|+1) \\
&= -(0.9)\left(\frac{1}{2}\right)\log 0.9 - (0.1)\left(\frac{1}{2}\right)\log 0.1 \\
&\quad +(0.1)\left(\frac{1}{2}\right)\log 0.1 - (0.9)\left(\frac{1}{2}\right)\log 0.9.
\end{aligned}$$

Thus $H(X_2|X_1) = -0.9\log 0.9 - 0.1\log 0.1 = h_b(0.1)$.

Now to compute $H(X_i|X_{i-1}, X_{i-2})$, first realize that since the dog started at 0, then at time i, the maximum and minimum positions it can be at are $+i$ and $-i$ respectively. Hence, $X_{i-1} \in \{-(i-1), -i+2, -i+3, \ldots, i-2, i-1\}$. Thus:

$$\begin{aligned}
H(X_i|X_{i-1}, X_{i-2}) &= -\sum_{x_{i-1}}\sum_{x_i}\sum_{x_{i-2}} p_{X_i|X_{i-1},X_{i-2}}(x_i|x_{i-1}, x_{i-2}) \\
&\quad p_{X_{i-1}X_{i-2}}(x_{i-1}, x_{i-2})\log p(x_i|x_{i-1}, x_{i-2}) \\
&= -\sum_{k=-(i-1)}^{i-1}\ \sum_{l=k-1,k+1}\ \sum_{j=k-1,k+1} p_{X_i|X_{i-1},X_{i-2}}(l|k,j) \\
&\quad p_{X_{i-1}X_{i-2}}(k,j)\log p_{X_i|X_{i-1},X_{i-2}}(l|k,j).
\end{aligned}$$

Now observe that

$$\begin{aligned}
p_{X_i|X_{i-1},X_{i-2}}(k+1|k,k-1) &= 0.9 \\
p_{X_i|X_{i-1},X_{i-2}}(k-1|k,k-1) &= 0.1 \quad (*) \\
p_{X_i|X_{i-1},X_{i-2}}(k+1|k,k+1) &= 0.1 \quad (*) \\
p_{X_i|X_{i-1},X_{i-2}}(k-1|k,k+1) &= 0.9
\end{aligned}$$

where $(*)$ denotes the cases of reversal in direction. Now using the following shorthand notation

$$p_{X_i|X_{i-1},X_{i-2}}(\cdot|\cdot,\cdot) = p(\cdot|\cdot,\cdot), \qquad \text{and} \qquad p_{X_{i-1}X_{i-2}}(\cdot,\cdot) = p(\cdot,\cdot),$$

we get

$$\begin{aligned}
H(X_i|X_{i-1},X_{i-2}) &= -\sum_{k=-(i-1)}^{i-1} [p(k-1|k,k-1)p(k,k-1)\log p(k-1|k,k-1) \\
&+ \; p(k-1|k,k+1)p(k,k+1)\log p(k-1|k,k+1) \\
&+ \; p(k+1|k,k-1)p(k,k-1)\log p(k+1|k,k-1) \\
&+ \; p(k+1|k,k+1)p(k,k+1)\log p(k+1|k,k+1)] \\
&= \sum_{k=-(i-1)}^{i-1} \quad p_{X_{i-1}X_{i-2}}(k,k-1)h_b(0.1) \\
&\qquad\qquad + \; p_{X_{i-1}X_{i-2}}(k,k+1)h_b(0.1)] \\
&= h_b(0.1)[\underbrace{\sum_{k=-(i-1)}^{i-1} [p_{X_{i-1}X_{i-2}}(k,k-1) + p_{X_{i-1}X_{i-2}}(k,k+1)]}_{=\,1\,(**)}.
\end{aligned}$$

$(**)$ is true since these are all the possible values that X_{i-1} and X_{i-2} can jointly have. Thus

$$\begin{aligned}
H(X_i|X_{i-1},X_{i-2}) &= h_b(0.1). \\
H(X_1,\ldots,X_n) &= 1 + h_b(0.1) + (n-2)h_b(0.1) \\
&= 1 + (n-1)h_b(0.1) \\
&= 1 + 0.469(n-1).
\end{aligned}$$

(b) $H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1,\ldots,X_n) = \lim_{n\to\infty} \frac{1+0.469(n-1)}{n} = 0.469$ bits/sample.

(c) The dog needs to take at least one step to establish the direction of his walk from which it ultimately reverses. Let K be the number of steps between reversals. K has a geometric distribution $P(K=k) = (0.9)^{k-1}(0.1), \quad k = 1,2,\ldots$. Thus

$$E[K] = \sum_{k=1}^{\infty} k(0.9)^{k-1}(0.1) = \frac{1}{0.1} = 10.$$

Problem B: *Binary Symmetric Markov Source*

(a) Solving $\pi = \pi Q$ where $\pi = [\pi_0, \pi_1]$ and

$$Q = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

yields $\pi = [1/2, 1/2]$. Thus need $P(X_1 = 0) = P(X_1 = 1) = 1/2$ to get a stationary Markov process.

(b) If $p = 1/2$, Thus, $P(X_n = 0|X_{n-1}) = P(X_n = 1|X_{n-1}) = \frac{1}{2} = P(X_n = 1) = P(X_n = 0)$. Thus, X_n's are i.i.d.

(c) If $p = 1$, $\{X_n\}$ is not irreducible and not ergodic.

(d) If $p = 0$, $\{X_n\}$ is irreducible $\Leftrightarrow$ ergodic.

(e) $H(\mathcal{X}) = H(X_2|X_1) = h_b(p)$, where $h_b(x) = -x \log x - (1-x)\log(1-x)$.

$\square$

Please let me know if you see a mistake in this solution.