Problem 2.10

(a) Define a new random variable as follows:

\[
E = \begin{cases} 
0 & \text{if } X = X_1, \\
1 & \text{if } X = X_2. 
\end{cases}
\]  

Then \( \Pr(E = 0) = \alpha = 1 - \Pr(E = 1) \) and \( H(E|X) = 0 \). We can get \( H(X) \) as

\[
H(X) = H(X, E) - H(E|X) = H(X, E) = H(E) + H(X|E) = h(\alpha) + \sum_{e \in \{0, 1\}} \Pr(E = e)H(X|E = e)
\]

\[
= h(\alpha) + \alpha H(X_1) + (1 - \alpha) H(X_2)
\]

where \( h(\alpha) \triangleq -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha) \).

(b) Start from Eqn.(6)

\[
H(X) = h(\alpha) + \alpha H(X_1) + (1 - \alpha) H(X_2)
\]

\[
= -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha) + \alpha H(X_1) + (1 - \alpha) H(X_2)
\]

\[
= -\alpha \log \frac{\alpha}{2H(X_1)} - (1 - \alpha) \log \frac{1 - \alpha}{2H(X_2)}
\]

\[
\leq -\log \frac{1}{2H(X_1) + 2H(X_2)}
\]

\[
= \log \left( 2H(X_1) + 2H(X_2) \right)
\]

where Eqn.(10) is due to log sum inequality. Then, Eqn.(11) yields

\[
2H(X) \leq 2H(X_1) + 2H(X_2)
\]

Problem 2.18

If \( A \) or \( B \) wins after \( Y = 5 \) turns, then the possible sequences of \( X \) are:

\[
BAAAA, AABAA, AAABA, AABAA, ABAAA, BABBB, BBABB, BBBBB
\]

i.e., there are in total \( 2 \times \binom{4}{1} \) sequences, and the probability of the event one of them happens is \( 1/2^5 \). Therefore, \( P(Y = 5) = 2 \times \binom{4}{1} \times \left( \frac{1}{2} \right)^5 \).
Then, we can calculate $H(X), H(Y), H(X|Y)$ and $H(Y|X)$:

$$H(X) = -\sum_{n=4}^{7} 2 \cdot \binom{n-1}{3} \left(\frac{1}{2}\right)^n \log \left(\frac{1}{2}\right)^n$$

(15)

$$\approx 5.813 \text{ (bits)}$$

(16)

$$H(Y) = -\sum_{n=4}^{7} 2 \cdot \binom{n-1}{3} \left(\frac{1}{2}\right)^n \log \left[ 2 \cdot \binom{n-1}{3} \left(\frac{1}{2}\right)^n \right]$$

(17)

$$\approx 1.924 \text{ (bits)}$$

(18)

$$H(Y|X) = 0$$

(19)

$$H(X|Y) = H(X) + H(Y|X) - H(Y) = 3.889 \text{ (bits)}$$

(20)

**Problem 2.26**

(a) Start from the Problem 2.13, i.e., $\forall x > 0$

$$\ln x \geq 1 - \frac{1}{x}$$

(21)

which yields

$$-\ln x = \ln \frac{1}{x} \geq 1 - x$$

(22)

i.e.,

$$\ln x \leq x - 1$$

(23)

where equality holds iff $x = 1$.

(b) If there exists $x$ such that $p(x) > 0$ and $q(x) = 0$, by the convention (Textbook, Page 19), $D(p||q) = \infty > 0$. Then we are done.

Otherwise, let $\mathcal{A} \triangleq \{ x \in \mathcal{X} : p(x) > 0 \}$. Based on this definition, $\forall x \in \mathcal{A}$, $\frac{q(x)}{p(x)}$ is well-defined and positive. Therefore, we can derive the upper bound $-D(p||q)$ as

$$-D(p||q) = \sum_{x \in \mathcal{X}} p(x) \ln \frac{q(x)}{p(x)}$$

(24)

$$= \sum_{x \in \mathcal{A}} p(x) \ln \frac{q(x)}{p(x)}$$

(25)

$$= \sum_{x \in \mathcal{A}} p(x) \ln \frac{q(x)}{p(x)}$$

(26)

$$\leq \sum_{x \in \mathcal{A}} p(x) \left[ \frac{q(x)}{p(x)} - 1 \right]$$

(27)

$$= \sum_{x \in \mathcal{A}} [q(x) - p(x)]$$

(28)

$$= \sum_{x \in \mathcal{A}} q(x) - \sum_{x \in \mathcal{X}} p(x)$$

(29)

$$= \sum_{x \in \mathcal{A}} q(x) - 1$$

(30)

$$\leq 0$$

(31)

(c) The inequality (27) holds iff $p(x) = q(x)$ for all $x \in \mathcal{A}$. Since $1 = \sum_{x \in \mathcal{A}} p(x) = \sum_{x \in \mathcal{X}} p(x) = \sum_{x \in \mathcal{X}} q(x)$, the condition $p(x) = q(x)$ for all $x \in \mathcal{A}$ implies that $q(x) = 0$ for all $x \in \mathcal{X}\setminus\mathcal{A}$, which is the sufficient and necessary condition when the inequality (31) holds. In sum, $p = q$. 

2
Problem 2.29

(1) Start from the left hand side (LHS)

\[ LHS = H(X|Z) + H(Y|X,Z) \geq H(X|Z) = RHS \]

where the equality holds iff \( Y \) is a deterministic function of \( X, Z \).

(2) Similarly, start from the left hand side (LHS)

\[ LHS = I(X;Z) + I(Y;Z|X) \geq I(X;Z) = RHS \]

where the equality holds iff \( Z \rightarrow X \rightarrow Y \) forms a Markov chain.

(3) Start from the left hand side (LHS)

\[ LHS = H(Z|X,Y) = H(Z|X) - I(Z;Y|X) \leq H(Z|X) = RHS \]

where the equality holds iff \( Z \rightarrow X \rightarrow Y \) forms a Markov chain.

(4) This is actually an equality.

\[ LHS = I(X,Y;Z) - I(Z;Y) = I(X;Z) + I(Y;Z|X) - I(Z;Y) = RHS \]

Problem 2.30

(a) The formal way is based on Lagrangian multiplier method, which is pretty straight-forward.

(b) Here we give another approach. Let \( q(n) = ab^n \) for \( n \geq 0 \). Then, by the nonnegative property of information divergence,

\[ D(p||q) = \sum_n p(n) \log \frac{p(n)}{q(n)} \geq 0 \]

which means

\[ H(X) = -\sum_n p(n) \log p(n) \leq -\sum_n p(n) \log q(n) = -\sum_n p(n) \log (ab^n) = -\log a - A \log b \]

which is clearly an upper bound (and is a constant). Equality holds iff \( p = q \) and \( q \) is a pmf.

The next step is to find such a pmf achieving the upper bound. By solving the following equation array,

\[ \begin{cases} 1 = \sum_n ab^n = \frac{a}{1-b} \\ A = \sum_n n ab^n = \frac{ab}{(1-b)^2} \end{cases} \]

which gives the optimal pmf \( p^*(n) = \frac{1}{1+A} \left( \frac{A}{1+A} \right)^n \) for \( n \geq 0 \) with \( \max H(X) = -\log \frac{1}{1+A} - A \log \frac{A}{1+A} \).
**Problem 3.3**

Let \( X_i \) be the proportion of piece at time \( t \). Note that the sequences \( \{X_i\} \) are i.i.d. with the distribution

\[
\Pr\left(X_i = \frac{2}{3}\right) = \frac{3}{4} \quad \text{and} \quad \Pr\left(X_i = \frac{3}{5}\right) = \frac{1}{4}
\] (46)

Then, after \( n \) cuts, the size of the case \( X \) is \( X = \prod_{i=1}^{n} X_i \). Take the log and use the law of large numbers

\[
\frac{1}{n} \log X = \frac{1}{n} \sum_{i=1}^{n} \log X_i
\] (47)

\[
\to \mathbb{E}[\log X_i]
\] (48)

\[
= \frac{3}{4} \log \frac{2}{3} + \frac{1}{4} \log \frac{3}{5}
\] (49)

i.e., the first order in exponent is \((\frac{3}{4} \log \frac{2}{3} + \frac{1}{4} \log \frac{3}{5})n\).

**Problem 3.4**

By AEP and the weak law of large numbers, both probabilities \( \Pr(X^n \in A^n) \) and \( \Pr(X^n \in B^n) \) go to the limit 1, i.e., for \( \epsilon' > 0 \), \( \Pr(X^n \in A^n) \geq 1 - \epsilon' \) and \( \Pr(X^n \in B^n) \geq 1 - \epsilon' \) for sufficient large \( n \). So (a) is true.

For (b),

\[
1 - \Pr(X^n \in A^n \cap B^n) = \Pr(X^n \in \overline{A^n} \cup \overline{B^n})
\leq \Pr(X^n \in \overline{A^n}) + \Pr(X^n \in \overline{B^n})
\leq \epsilon' + \epsilon' \triangleq \epsilon
\] (50)

which implies that \( \Pr(X^n \in A^n \cap B^n) \to 1 \).

For (c), for all \( n \)

\[
|A^n \cap B^n| \leq |A^n| \leq 2^{n(H + \epsilon)}
\] (53)

For (d), for sufficient large \( n \), we have Eqn.(52), i.e.,

\[
1 - \epsilon \leq \Pr(X^n \in A^n \cap B^n)
\leq \sum_{x^n \in A^n \cap B^n} \Pr(x^n)
\leq \sum_{x^n \in A^n \cap B^n} 2^{-n(H - \epsilon)}
= |A^n \cap B^n|2^{-n(H - \epsilon)}
\] (58)

which means

\[
|A^n \cap B^n| \geq (1 - \epsilon)2^{n(H - \epsilon)}
\] (59)

\[
\geq \frac{1}{2}2^{n(H - \epsilon)}
\] (60)

where the last inequality is simply choosing \( \epsilon < 1/2 \).

**Problem II**

Define \( p \triangleq P(X \neq Y) \) and \( r \triangleq |x||y| \). By Fano’s inequality,

\[
H(X|Y) \leq H(p) + p \log |x|
\] (61)

\[
H(Y|X) \leq H(p) + p \log |y|
\] (62)
one upper bound of $\rho(X,Y)$ is given as follows

$$\rho(X,Y) \leq 2H(p) + p \log r$$

$$= -2p \log p - 2(1 - p) \log(1 - p) + p \log r$$

$$= \frac{1}{\ln 2} \left[ -2p \ln p - 2(1 - p) \ln(1 - p) + p \ln r \right]$$

To find an explicit expression of $p$ such that $\rho(X,Y) < \epsilon$, we need to get the upper bounds for each part of Eqn.(65).

The third item $p \ln r$ is a linear function of $p$. The second item $-2(1 - p) \ln(1 - p)$ satisfies

$$2(1 - p) \ln \frac{1}{1 - p} \leq 2(1 - p) \left( \frac{1}{1 - p} - 1 \right) = 2p$$

For the first item $-2p \ln p$, we need the following inequality

$$-p \ln p \leq 2\sqrt{p}$$

There are several ways to prove this. One method based on Eqn.(23) is that for all $x > 0$

$$\ln x \leq x - 1$$

which is equivalent to say

$$\ln \sqrt{\frac{1}{x}} \leq \sqrt{\frac{1}{x}} - 1$$

i.e.,

$$\frac{1}{2} \ln \frac{1}{x} \leq \sqrt{\frac{1}{x}} - 1$$

which implies

$$x \ln \frac{1}{x} \leq 2x \left( \sqrt{\frac{1}{x}} - 1 \right) \leq 2\sqrt{x}$$

In sum, the Eqn.(65) can be

$$\rho(X,Y) = \frac{1}{\ln 2} \left[ -2p \ln p - 2(1 - p) \ln(1 - p) + p \ln r \right]$$

$$= \frac{1}{\ln 2} (4\sqrt{p} + 2p + p \ln r)$$

$$\leq \frac{1}{\ln 2} (4\sqrt{p} + 2\sqrt{p} + \ln r \sqrt{p})$$

$$= \frac{6 + \ln r}{\ln 2} \sqrt{p}$$

where $p < 1$ guarantees $p < \sqrt{p}$. If we choose

$$p \leq \left( \frac{\epsilon}{6 + \ln r} \right)^2$$

then $\rho(X,Y) \leq \epsilon$.

**Problem III**

Please note that the log function is the binary logarithm, which uses base 2.

**(a)** Let $p,q$ be any general pmfs defined on a finite set $\mathcal{X}$. Consider the set $M = \{x : p(x) \geq q(x)\}$. Then construct the Bernoulli pmfs $\tilde{p} = (p(M), 1 - p(M)), \tilde{q} = (q(M), 1 - q(M))$. 

Firstly, we have
\[
D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} = \sum_{x \in \mathcal{M}} p(x) \log \frac{p(x)}{q(x)} + \sum_{x \notin \mathcal{M}} p(x) \log \frac{p(x)}{q(x)} \geq p(M) \log \frac{p(M)}{q(M)} + [1 - p(M)] \log \frac{1 - p(M)}{1 - q(M)} = D(\tilde{p}||\tilde{q})
\]
where the inequality is due to log sum inequality.

On the other hand,
\[
d(p,q) = \sum_{x \in \mathcal{X}} |p(x) - q(x)| = \sum_{x \in \mathcal{M}} |p(x) - q(x)| + \sum_{x \notin \mathcal{M}} |p(x) - q(x)| = d(\tilde{p}, \tilde{q})
\]

Therefore, if we could prove
\[
D(\tilde{p}||\tilde{q}) \geq \frac{1}{2 \ln 2} d^2(\tilde{p}, \tilde{q})
\]
then
\[
D(p||q) \geq D(\tilde{p}||\tilde{q}) \geq \frac{1}{2 \ln 2} d^2(\tilde{p}, \tilde{q}) = \frac{1}{2 \ln 2} d^2(p,q)
\]
The equivalent form of Eqn.(84) is shown as follows
\[
p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} - \frac{2}{\ln 2} (p - q)^2 \geq 0
\]
where for convenience we change the notation \((\tilde{p}, \tilde{q}) \rightarrow (p,q)\).

It suffices to prove that for any \(0 \leq p, q \leq 1\) Eqn.(86) always satisfies. Fix \(p\), denote
\[
f_p(q) \triangleq p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} - \frac{2}{\ln 2} (p - q)^2
\]
and take the derivative
\[
f'_p(q) = \frac{1}{\ln 2} \left[ \frac{p}{q} + \frac{1 - p}{1 - q} + 4(p - q) \right] = \frac{q - p}{\ln 2} \left[ \frac{1}{q(1 - q)} - 4 \right] = \frac{q - p}{\ln 2} \Delta
\]
Since \(\frac{1}{q(1 - q)} \geq 4\), \(\Delta \triangleq \left[ \frac{1}{q(1 - q)} - 4 \right] \geq 0\), which implies that
\[
\begin{cases}
    f'_p(q) = 0 & \text{if } q = p \\
    f'_p(q) \geq 0 & \text{if } q > p \\
    f'_p(q) \leq 0 & \text{if } q < p
\end{cases}
\]
i.e., \(f_p(q)|_{q=p} = 0\) is a minimum point, which is valid for all \(p\). Therefore, Eqn.(86) is true for all \(0 \leq p, q \leq 1\).

(b) To find the \(p, q\) such that \(\left| \frac{D(p||q)}{d^2(p,q)} - \frac{1}{2 \ln 2} \right| < \epsilon\) for small \(\epsilon > 0\), it suffices to consider Bernoulli pmfs.
Note that, for any \( p \neq q \), \( d(p, q) \neq 0 \). By Eqn.(84), it is equivalent to solve

\[
0 \leq \frac{D(p||q)}{d^2(p, q)} - \frac{1}{2\ln 2} < \epsilon
\]  

(92)

After choosing \( q = \frac{1}{2} \) and \( p = \frac{1}{2} + \delta \), (\( \delta > 0 \)), it is easy to check that

\[
\lim_{\delta \to 0} \frac{D(p||q)}{d^2(p, q)} - \frac{1}{2\ln 2} = \lim_{\delta \to 0} \frac{1 - h \left( \frac{1}{2} + \delta \right)}{4\delta^2} - \frac{1}{2\ln 2} = 0
\]  

(93)

The idea is that, since \( \lim_{\delta \to 0} \left[ 1 - h \left( \frac{1}{2} + \delta \right) \right] = 0 \), we can use l’Hospital’s rule. The limit of the first item \( \frac{1 - h \left( \frac{1}{2} + \delta \right)}{4\delta^2} \) is exactly \( \frac{1}{\ln 2} \). This means that we can choose sufficient small \( \delta \) to make \( \left| \frac{D(p||q)}{d^2(p, q)} - \frac{1}{2\ln 2} \right| < \epsilon \).