Below $h(p)$ denote the binary entropy function.

**Question 1**

- The score function is defined as follows

$$ V(X) = \frac{\partial}{\partial \theta} \ln f(X; \theta) $$

$$ \ln(f(x; \theta)) = \ln(\theta) - \theta x $$

$$ V(X) = \frac{1}{\theta} - X. $$

- The Fisher Information is given as

$$ J(\theta) = E_{\theta}[V^2] $$

$$ J(\theta) = \int (\frac{1}{\theta} - x)^2 f(x; \theta) dx $$

$$ J(\theta) = \frac{1}{\theta^2} $$

Cramer-Rao bound on the unbiased estimator is

$$ E[(T - \theta)^2] \geq \frac{1}{J(\theta)} = \theta^2 $$

- Given iid samples, then

$$ J_n(\theta) = nJ(\theta) = \frac{n}{\theta^2} $$

Thus the Cramer Rao bound is

$$ E[(T - \theta)^2] \geq \frac{\theta^2}{n} $$

- This is a trick question as it is quite difficult to come up with an unbiased estimator for $\theta$ in this problem. Instead, we design an estimator for $\mu = \frac{1}{\theta}$.

The estimator for $\mu$ that I propose is

$$ \hat{\mu} \equiv \mu(X_1, \ldots, X_n) = n \min\{X_1, \ldots, X_n\}. $$
Since,
\[ \Pr(\hat{\mu} \leq t) = 1 - \Pr(\min\{X_1, \ldots, X_n\} \geq t/n) = 1 - (e^{-\theta t/n})^n = 1 - e^{-\theta t}, \]
hence $\hat{\mu}$ is an exponential distribution with mean $\mu$. Therefore this is an unbiased estimator of $\mu$.

Our overall estimator is $T(X_1, \ldots, X_n) = \frac{1}{\min(X_1, \ldots, X_n)}$. This estimator does not meet the Cramer-Rao bound. However the estimator for the $\mu$ does.

**Question 2**

When $a = 0$ the capacity of the channel is obviously 1 (since $Y = X$). We consider the other cases below.

For this channel,
\[ I(X; Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x). \]

Now,
\[ H(Y|X = 0) = 1 = H(Y|X = 1). \]

Hence,
\[ I(X; Y) = H(Y) - 1. \]

Assume, $P(X = 1) = q$ and $P(X = 0) = 1 - q$. Now consider the following cases.

- **a=1**: Here $Y = \{0, 1, 2\}$. We have $P(Y = 0) = \frac{1}{2}(1 - q)$, $P(Y = 1) = \frac{1}{2}q + \frac{1}{2}(1 - q) = \frac{1}{2}$ and $P(Y = 2) = \frac{a}{2}$. Therefore, $H(Y)$ will be maximized when $1 - q = q$ or $q = \frac{1}{2}$. The maximum value of $H(Y)$ is $H(1/4, 1/2, 1/4) = 2 \times 1/4 \times 2 + 1/2 = \frac{3}{2}$. Hence the capacity is $\frac{3}{2} - 1 = \frac{1}{2}$.

- **a=-1**: The calculations are exactly same as above.

- **Any other a**: Here $Y = \{0, 1, a, 1 + a\}$. We have, $P(Y = 0) = \frac{1 - q}{2}$, $P(Y = 1) = \frac{q}{2}$, $P(Y = a) = \frac{1 - q}{2}$, $P(Y = 1 + a) = \frac{a}{2}$. Clearly the maximum value that $H(Y)$ can have is 2 and that happens when $q = \frac{1}{2}$. So the capacity is $2 - 1 = 1$.

**Question 3**

Similar to the Question 2 we can write
\[ C = \max_{p(x)}[H(Y) - \sum_x p(x)H(Y|X = x)] \]

Assume that the input distribution is given by,
\[ P(X = 0) = 1 - q; P(X = 1) = q. \]
Then
\[ P(Y = 0) = 1 - q + qp; \quad P(Y = 1) = q(1 - p). \]
We have \( H(Y|X = 0) = 0 \) and \( H(Y|X = 1) = h(p) \), hence,
\[ H(Y|X) = qh(p). \]
Therefore,
\[ I(X; Y) = h(q(1 - p)) - qh(p). \]

Differentiating this and equating to 0 we have,
\[ (1 - p) \log \frac{1 - q(1 - p)}{q(1 - p)} - h(p) = 0 \]
or
\[ 2^{h(p)} = \frac{1 - q(1 - p)}{q(1 - p)} = \frac{1}{q(1 - p)} - 1 \]
or
\[ q(1 - p) = \frac{1}{1 + 2^{h(p)}.} \]

This gives the capacity to be,
\[
\begin{align*}
&h\left(\frac{1}{1 + 2^{\frac{h(p)}{1 - p}}}\right) - \frac{1}{(1 - p)(1 + 2^{\frac{h(p)}{1 - p}})}h(p) \\
&= \frac{1}{1 + 2^{\frac{h(p)}{1 - p}}} \log(1 + 2^{\frac{h(p)}{1 - p}}) + \frac{2^{\frac{h(p)}{1 - p}}}{1 + 2^{\frac{h(p)}{1 - p}}} \left(\log(1 + 2^{\frac{h(p)}{1 - p}}) - \frac{h(p)}{1 - p}\right) - \frac{h(p)}{1 - p} \frac{1}{1 + 2^{\frac{h(p)}{1 - p}}}
\end{align*}
\]
\[ = \log(1 + 2^{\frac{h(p)}{1 - p}}). \]

**Question 4**

- All pigeons transfer information without any error. Number of pigeons sent (and received) every hour is 12. Each pigeon carries 8 bits. So total number of bits sent in an hour is 96 bits. All these bits reach as they were sent.

- We can model this as an erasure channel. Here, each combination of 8 bits is a new symbol. So a total of 256 input symbols. The output alphabet is made of 257 symbols. An extra symbol for erasure. And this erasure symbol occurs with probability \( \alpha \). So we know that
\[
I(X; Y) = H(Y) - h(\alpha) \quad (11)
\]

Now to maximize \( H(Y) \), we can't directly consider it as a uniform distribution over all symbols of \( Y \). Since, the erasure channel probability \( \alpha \) of erasing is independent of the input distribution. Let \( E \) be the erasure event.
\[
H(Y) = H(E) + H(Y|E) = h(\alpha) + (1 - \alpha)H(Y|E = 0) \quad (12)
\]
Now we can surely maximize $H(Y|E = 0)$ by considering a uniform distribution over the input symbols. That is $\log 256 = 8$. Thus

$$C = \max_{p(x)} I(X; Y) = (1 - \alpha)8. \quad (13)$$

Hence the capacity in bits per hour will be $96(1 - \alpha)$.

- For this case we have a symmetric channel with 256 inputs and 256 outputs. An input symbol $x'$ maps to the same output symbol $x'$ with a probability $(1 - \alpha) + \frac{\alpha}{256}$ and it maps to any other output symbol with a probability $\frac{\alpha}{256}$.

$$I(X; Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \sum_x p(x)H(Y|X = x)$$

$$= H(Y) - \sum_x p(x)H(1 - \alpha + \frac{\alpha}{256}, \frac{\alpha}{256}, \frac{\alpha}{256}, \ldots, \frac{\alpha}{256})$$

$$= H(Y) - H(1 - \alpha + \frac{\alpha}{256}, \frac{\alpha}{256}, \frac{\alpha}{256}, \ldots, \frac{\alpha}{256})$$

$$= H(Y) + (1 - \alpha + \frac{\alpha}{256}) \log(1 - \alpha + \frac{\alpha}{256}) + \frac{255\alpha}{256} \log \frac{256}{256}$$

$$= H(Y) + (1 - \frac{255\alpha}{256}) \log(1 - \frac{255\alpha}{256}) + \frac{255\alpha}{256} \log \frac{256}{256}.$$

Now this mutual information will be maximum when the distribution of output symbols is uniform, which is possible in this case (just by taking the input to be uniform). Thus,

$$C = 8 + (1 - \frac{255\alpha}{256}) \log(1 - \frac{255\alpha}{256}) + \frac{255\alpha}{256} \log \frac{256}{256}. \quad (14)$$

**Question 5**

For this problem, $h(X)$ denote the differential entropy of a continuous random variable $X$. We have,

$$I(X; Y) = h(Y) - h(Y|X)$$

$$= h(Y) - h(Z + X|X)$$

$$= h(Y) - h(Z|X)$$

$$= h(Y) - h(Z).$$

We know that,

$$f_Z(z) = \frac{1}{\mu} e^{-\frac{z}{\mu}}, \quad z \geq 0.$$

Hence,

$$h(Z) = -\int_0^\infty f_Z(z)(- \ln \mu - \frac{z}{\mu})dz = \ln \mu + 1.$$
We further have,
\[ EY = EX + EZ \leq \mu + \lambda. \]

We claim that, among the distributions with nonnegative support and bounded mean, exponential distribution has the highest entropy. Consider \( f \) to be an exponential distribution of mean \( \theta \) and \( g \) be any other distribution of mean \( \theta \). Then,
\[
\begin{align*}
    h(f) - h(g) &= -\int_0^\infty f(x)(-\ln \frac{x}{\theta}) dx + \int_0^\infty g(x) \ln g(x) dx \\
    &= -\int_0^\infty g(x)(-\ln \frac{x}{\theta}) dx + \int_0^\infty g(x) \ln g(x) dx \\
    &= D(g\|f) \geq 0.
\end{align*}
\]

This means,
\[ h(Y) \leq \ln(\mu + \lambda) + 1. \]

Hence,
\[ I(X;Y) \leq \ln(\mu + \lambda) - \ln \mu = \ln(1 + \frac{\lambda}{\mu}). \]

Since it is always possible to find out a distribution of \( X \) where \( X = Y - Z \) where \( Y \) and \( Z \) are exponential distributions of mean \( \mu + \lambda \) and \( \mu \) respectively, the capacity of the channel is
\[ \ln(1 + \frac{\lambda}{\mu}). \]

**Question 6**

- Parity check matrix has dimensions \( n \times (n-k) \). In this case \( n=15 \), and \( (n-k) = 4 \). So \( k =11 \). There exists a set of 3 linearly dependent columns, but any two columns are linearly independent. Hence minimum distance \( d = 3 \).

- A Generator matrix is given by

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\]
A parity check matrix is given by,
\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\end{bmatrix}
\]

Multiplying this with \(x = [000000z_1z_20000111]^T\) we have,
\[z_1[0111]^T + z_2[1000]^T = [1100]^T.\]

No matter what \(z_1\) or \(z_2\) you choose, this equation is not satisfied, hence there must be an error.

Let us look at the candidate codewords:

- **Case 1:** \(z_1 = 0\) and \(z_2 = 0\). Here \(Hx = [1100]^T\), which is the 12th column of \(H\). Hence the decoded codeword would be, \([00000000001111]^T\).
- **Case 2:** \(z_1 = 0\) and \(z_2 = 1\). Here \(Hx = [0100]^T\), which is the 4th column of \(H\). Hence the decoded codeword would be, \([00010001000011]^T\).
- **Case 3:** \(z_1 = 1\) and \(z_2 = 0\). Here \(Hx = [1011]^T\), which is the 11th column of \(H\). Hence the decoded codeword would be, \([00000100011111]^T\).
- **Case 4:** \(z_1 = 1\) and \(z_2 = 0\). Here \(Hx = [0011]^T\), which is the 3rd column of \(H\). Hence the decoded codeword would be, \([00100011001111]^T\).

**Question 7**

- The length \(n\) of the code is 6. The dimension \(k\) of the code is 3. Hence, the number of code words is 8.
- A parity check matrix is given as follows:
\[
H = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
\end{bmatrix}
\]  
(15)

- The distance \(d\) of the code is equal to 3. Since, any 2 columns of \(H\) are linearly independent. Since \(d = 3\), the code can correct any 1 error and any 2 erasures.
- To decode the received word: 111000, we multiply it with the parity check matrix:
\[
H \times \begin{bmatrix}
1 \\
1 \\
0 \\
0 \\
\end{bmatrix} = \begin{bmatrix}
1 \\
0 \\
0 \\
\end{bmatrix}
\]  
(16)

Therefore the 4th bit has been flipped and the decoded codeword is 111100.