1

(a)

true:

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

If $I(X;Y) = 0$ then $H(X) = H(X|Y)$. We can write:

$$I(X;Y) = D(\ P_{x,y}(x,y)\ ||\ P_x(x)P_y(y)\ ) = 0$$

$D(Q||P) = 0$ iff $P_x(x) = Q_x(x)\ \forall x$, therefore $P_{x,y}(x,y) = P_x(x)P_y(y)$ for every x, y and as result $X \perp Y$.

(b)

false: If $p(x|y) \perp p(y|z)$ then $H(X|Y)$ and $(Y|Z)$ have any ratio between them.

(c)

true:
Using the concave property of the divergence function:

$$D\left(\lambda P + (1-\lambda)Q\ ||\ Q\right) \leq \lambda D(P\ ||\ Q) + (1-\lambda)D(Q\ ||\ Q)$$

Assigning $\lambda = \frac{1}{2}$, and since $D(Q||Q) = 0$:

$$D\left(\frac{1}{2}P + \frac{1}{2}Q\ ||\ Q\right) \leq \frac{1}{2}D(P||Q)$$

(d)

false:
We have proven the inequality $H(g(Z)) \leq H(Z)$ on homework. This time Z is a random variable with the joint distribution $P_{x,y}$. Therefore:

$$H(X+Y) \leq H(X,Y)$$

(e)

true:
Note: In general, $I(X;Y|Z)$ can be larger than $I(X;Y)$ and therefore $I(X;Y) - I(X;Y|Z)$ can be less then zero.

$$|I(X;Y) - I(X;Y|Z)| \quad = \quad \max\left\{[I(X;Y) - I(X;Y|Z)], [I(X;Y|Z) - I(X;Y)]\right\}$$

The first expression is:

$$\begin{aligned}
I(X;Y) - I(X;Y|Z) &= H(X) - H(X|Y) - [H(X|Z) - H(X|Y,Z)] \\
&= \underbrace{H(X) - H(X|Z)}_{I(X;Z)} - \underbrace{[H(X|Y) - H(X|Y,Z)]}_{\geq 0} \\
&\leq I(X;Z) \\
&= H(Z) - \underbrace{H(Z|X)}_{\geq 0} \\
&\leq H(Z)
\end{aligned}$$

The second expression is:

$$\begin{aligned}
I(X;Y|Z) - I(X;Y) &= H(X|Z) - H(X|Y,Z) - [H(X) - H(X|Y)] \\
&= H(X|Y) - H(X|Y,Z) - [H(X) - H(X|Z)] \\
&= I(X;Z|Y) - I(X;Z) \\
&\leq I(X;Z|Y) \\
&= H(Z|Y) - H(Z|X,Y) \\
&\leq H(Z|Y) \\
&\leq H(Z)
\end{aligned}$$

Therefore

$$|I(X;Y) - I(X;Y|Z)| \quad \leq \quad \max\{H(Z), H(Z)\} = H(Z)$$

(f)

false:
We know that $\frac{1}{n}\log|A_n| \geq H(X) - \varepsilon$ for n sufficiently large (*theorem 3.3.1* in the text book and as proved in class). Since $\lim_{n\to\infty} \Pr(A_n) = 1$ and $\lim_{n\to\infty} \Pr(B_n) = 1$ we can say that also $\lim_{n\to\infty} \Pr(A_n \cap B_n) = 1$ (it was also shown in class) and therefore:

$$\lim_{n\to\infty} \frac{1}{n} \log|A_n \cap B_n| \geq H(X) - \varepsilon$$

But since ε is as small as we like, we cannot say that:

$$\lim_{n\to\infty} \frac{1}{n} \log|A_n \cap B_n| < H(X)$$

(g)

false:
Assuming that the file is already optimally compressed, it cannot be compressed any further. Also, if the entropy rate of the bits in the file is 1 for some reason, it cannot be compressed.

For example, if the bits in the file are *Bernoulli*$(\frac{1}{2})$ distributed, the file cannot be compressed anymore.

(h)

true:
Using *theorem 8.4.1* and *example 8.5.1* in the text book:

$$h(X_1, X_2, \ldots, X_2) = \frac{1}{2} \log[(2\pi e)^n |K|]$$

Using the identity:

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

Assigning:

$$h(X, Y) = \frac{1}{2} \log[(2\pi e)^2 (\sigma^4 - \sigma^4 \rho^2)]$$

And:

$$h(Y) = \frac{1}{2} \log(2\pi e \sigma^2)$$

Therefore:

$$h(X, Y) - h(Y) = \frac{1}{2} \log[2\pi e \sigma^2 (1 - \rho^2)]$$

(i)

false:
Using the answer of the last question:

$$h(Y|X) = \frac{1}{2} \log[2\pi e \sigma^2 (1 - \rho^2)]$$

If $\sigma^2 \leq 2\pi e$ then $\log[2\pi e \sigma^2 (1 - \rho^2)] \leq 0$, and:

$$h(Y|X) \leq 0$$

(j)

true:
Increasing the distortion allows rate reduction.

(k)

true:

$$R(D) \leq I(X; \hat{X}) = H(X) - H(X|\hat{X}) = H(\hat{X}) - H(\hat{X}|X)$$

Therefore $R(D) \leq H(X)$ and $R(D) \leq H(\hat{X})$. And we can say that:

$$R(D) \leq \min(H(X), H(\hat{X}))$$

Using *theorem 2.6.4* ($H(X) \leq \log|\mathcal{X}|$):

$$R(D) \leq \min(\log|\mathcal{X}|, \log|\hat{\mathcal{X}}|)$$

Since log is a non descending function:

$$R(D) \leq \log(\min(|\mathcal{X}|, |\hat{\mathcal{X}}|))$$

2

(a)

Huffman code:

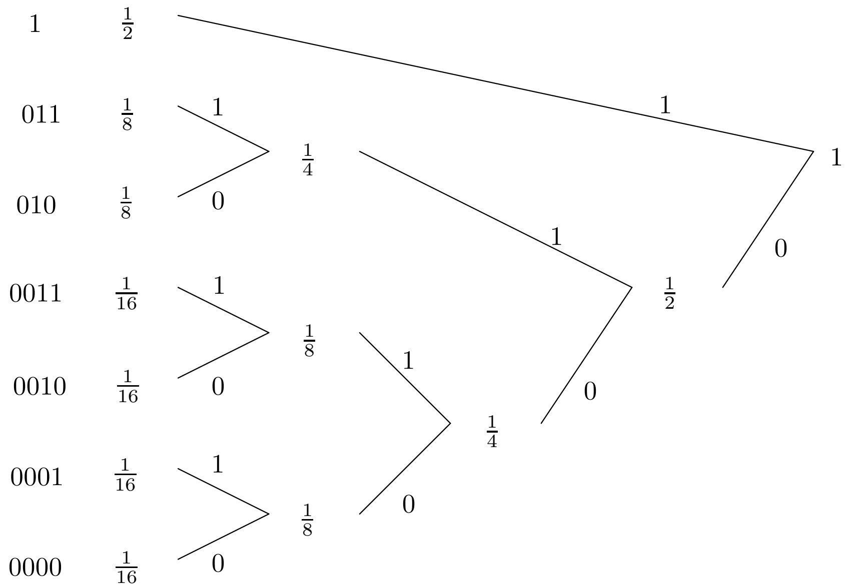

Figure 1: Huffman

(b)

Huffman code is optimal code and achieves the entropy for dyadic distribution. If the distribution of the digits is not $Bernoulli(\frac{1}{2})$ you can compress it further. The binary digits of the data would be equally distributed after applying the Huffman code and therefore $p_0 = p_1 = \frac{1}{2}$.

The expected length would be:

$$E[l] = \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 4 + \frac{1}{16} \cdot 4 + \frac{1}{16} \cdot 4 + \frac{1}{16} \cdot 4 = 2.25$$

Therefore, the expected length of 1000 symbols would be 2250 bits.

3

(a)

$$Y = h_1X_1 + h_2X_2 + Z$$

The mutual information is:

$$\begin{aligned} I(X_1, X_2; Y) &= h(Y) - h(Y|X_1, X_2) \\ &= h(Y) - h(Z) \end{aligned}$$

Since $h(z)$ is constant, we need to find the maximum of $h(Y)$, the second moment of Y is:

$$\begin{aligned} E[Y^2] &= E[(h_1X_1 + h_2X_2 + Z)^2] \\ &\overset{(i)}{=} E[(h_1X_1 + h_2X_2)^2] + E[Z^2] \\ &= h_1^2[X_1^2] + h_2^2[X_2^2] + 2h_1h_2E[X_1X_2] + \sigma_Z^2 \\ &\leq h_1^2P_1 + h_2^2P_2 + 2h_1h_2E[X_1X_2] + \sigma_Z^2 \\ &\overset{(ii)}{\leq} h_1^2P_1 + h_2^2P_2 + 2h_1h_2\sqrt{E[X_1^2]E[X_2^2]} + \sigma_Z^2 \\ &\leq h_1^2P_1 + h_2^2P_2 + 2h_1h_2\sqrt{P_1P_2} + \sigma_Z^2 \\ &= (h_1\sqrt{P_1} + h_2\sqrt{P_2})^2 + \sigma_Z^2 \end{aligned}$$

(i) - Z is independent of X_1, X_2.

(ii) - Cauchy-Schwarz inequality. Where $X_1 = \alpha X_2$, $\binom{X_1}{X_2} \sim N(0, K)$ and $K = \begin{pmatrix} P_1 & \sqrt{P_1P_2} \\ \sqrt{P_1P_2} & P_2 \end{pmatrix}$ will result with equality and bring the mutual information to a maximum.

Therefore, the mutual information is bounded by:

$$I(X_1, X_2; Y) \leq \frac{1}{2}\log\left(1 + \frac{(h_1\sqrt{P_1} + h_2\sqrt{P_2})^2}{\sigma_Z^2}\right)$$

(b)

The capacity of the system is:

$$C = \max_{P_{x_1,x_2}} I(X_1, X_2; Y) = \frac{1}{2}\log\left(1 + \frac{(h_1\sqrt{P_1} + h_2\sqrt{P_2})^2}{\sigma_Z^2}\right)$$

(c)

i

For $h_1 = 1$ and $h_2 = 1$ the capacity of the system would be:

$$
\begin{aligned}
C &= \frac{1}{2}\log\left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{\sigma_Z^2}\right) \\
&= \frac{1}{2}\log\left(1 + \frac{P_1 + 2\sqrt{P_1 P_2} + P_2}{\sigma_Z^2}\right)
\end{aligned}
$$

ii

For $h_1 = 1$ and $h_2 = 0$ the capacity of the system would be:

$$
C = \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_Z^2}\right)
$$

iii

For $h_1 = 0$ and $h_2 = 0$ the capacity of the system would be:

$$
C = \frac{1}{2}\log(1) = 0
$$

We can see that having 2 channels both increase the signal level and provides redundancy.

4

(a)

Since the noise is not know to both sides, the total noise is $\sigma_1^2 + \sigma_2^2$ and the capacity is:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_1^2 + \sigma_2^2}\right)$$

(b)+(C)

Once Z_2 is known to the receiver, we can add a subtraction unit in the decoder that subtract Z_2 and therefore the noise is only Z_1. And the capacity is:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_1^2}\right)$$