Handout 20 Information Theory and Coding
Solutions to Homework 8 Nov. 12, 2024

PROBLEM 1. The assertion is clearly true with $n = 1$. To complete the proof by induction we need to show that the cascade of a BSC with parameter $q = \frac{1}{2}(1-(1-2p)^n)$ with a BSC with parameter p is equivalent to a BSC with parameter $\frac{1}{2}(1-(1-2p)^{n+1})$. To do so, observe that for a cascade of a BSC with parameter q and a BSC with parameter p, when a bit is sent, the opposite bit will be received if exactly one of the channels makes a flip, and this happens with probability $(1-q)p + (1-p)q$. Thus, the cascade is equivalent to a BSC with this parameter. For $q = \frac{1}{2}(1-(1-2p)^n)$,

$$(1-q)p + (1-p)q = \frac{1}{2}(1+(1-2p)^n)p + \frac{1}{2}(1-(1-2p)^n)(1-p) = \frac{1}{2}(1-(1-2p)^{n+1}),$$

and the assertion is proved.

Alternate proof: the cascade makes flips the incoming bit if an odd number of the elements of the cascade flip. Thus the cascade is equivalent to a BSC with parameter

$$a = \sum_{k:k \text{ odd}} \binom{n}{k} p^k (1-p)^{n-k}.$$

Let $b = \sum_{k:k \text{ even}} \binom{n}{k} p^k (1-p)^{n-k}$. Observe that

$$a + b = \sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1,$$

and

$$-a + b = \sum_k \binom{n}{k} (-p)^k (1-p)^{n-k} = (-p + 1 - p)^n = (1-2p)^n.$$

Subtracting the two equalities and dividing by two, we get $a = \frac{1}{2}(1+(1-2p)^n)$.

PROBLEM 2. Let $P'_{X,Y}(x,y) = P_{Y|X}(y|x)Q'(x)$, $P'_Y(y) = \sum_{x\in\mathcal{X}} P'_{X,Y}(x,y)$ and $P_Y(y) = \sum_{x\in\mathcal{X}} P_{Y|X}(y|x)Q(x)$. We then have for any Q'

$$\begin{aligned}
&\sum_{x\in\mathcal{X}} Q'(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) \log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_{Y|X}(y|x')Q(x')}\right) - I(Q') \\
&= E_{P'_{X,Y}} \log \frac{P_{Y|X}}{P_Y} - I(Q') \\
&= E_{P'_{X,Y}} \left(\log \frac{P_{Y|X}}{P_Y} - \log \frac{P'_{X,Y}}{Q'_X P'_Y}\right) \\
&= E_{P'_{X,Y}} \log \frac{P'_Y}{P_Y} \\
&= E_{P'_Y} \log \frac{P'_Y}{P_Y} \\
&= D(P'_Y||P_Y) \geq 0
\end{aligned}$$

with equality if and only if $Q' = Q$. To prove (b), notice in the upper bound of part (a), that the inner summation is a function of x and that the outer summation is an average of this function with respect to the distribution $Q'(x)$. The average of a function is upper bounded by the maximum value that the function takes, and hence (b) follows.

PROBLEM 3. (a) By the chain rule

$$I(U,T;V) = I(U;V) + I(T;V|U) = I(U;V),$$

since $I(T;V|U) = 0$ from the Markov property. Also,

$$I(U,T;V) = I(T;V) + I(U;V|T) \geq I(U;V|T),$$

from the non-negativity of the mutual information. These together imply that $I(U;V) \geq I(U;V|T)$.

(b)

$$I(X;Y|W) = \Pr\{W=1\}I(X;Y|W=1) + \Pr\{W=2\}I(X;Y|W=2)$$

Conditional on $W=k$, the distribution of (X,Y) is $p_k(x)p(y|x)$, thus

$$I(X;Y|W) = \lambda I_1 + (1-\lambda)I_2.$$

(c) We obtain $p(x)$ by summing $p(w,x,y)$ over y and w. This gives

$$p(x) = \lambda p_1(x) + (1-\lambda)p_2(x).$$

(d) Note that

$$p(w,x,y) = p(w)p(x|w)p(y|x),$$

that is Y is independent of W when X is given. Thus from (a)

$$I(X;Y) \geq I(X;Y|W). \tag{1}$$

Letting $f(p_X)$ denote the value of $I(X;Y)$ as a function of the distribution of X we can rewrite (1) as

$$f(\lambda p_1 + (1-\lambda)p_2) \geq \lambda f(p_1) + (1-\lambda)f(p_2),$$

which says that f is concave.

PROBLEM 4. Since X and Z are both in the interval $[-1,1]$, their sum $X+Z$ lies in the interval $[-2,+2]$. If we could *choose* the distribution of $X+Z$ as we wished (without the constraint that it has to be the sum of two independent random variables, one of which is uniform) we would have chosen it to be uniform on the interval $[-2,+2]$ to have the largest entropy. Observe now that if we choose X as the random variable that equals $+1$ with probability $1/2$ and -1 with probability $1/2$, then $X+Z$ is uniform in $[-2,+2]$ and thus this distribution maximizes the entropy. An alternate derivation is as follows: note that since X and Z are independent, the moment generating functions of the random variables involved satisfy $E[e^{s(X+Z)}] = E[e^{sX}]E[e^{sZ}]$. Now, we know that $E[e^{sZ}] = \int e^{sz} f_Z(z)\,dz = \int_{-1}^{+1} \frac{1}{2} e^{sz}\,dz = [e^s - e^{-s}]/(2s)$. Similarly, if we want $X+Z$ to be uniform on $[-2,2]$, we can compute $E[e^{s(X+Z)}] = [e^{2s} - e^{-2s}]/(4s)$. This then requires $E[e^{sX}] = \frac{1}{2}[e^{2s} - e^{-2s}]/[e^s - e^{-s}] = \frac{1}{2}[e^s + e^{-s}]$ which is the moment generating function of a random variable which takes on the values $+1$ and -1, each with probability $1/2$.

Similarly, under the constraint XZ lies in the interval $[-1,+1]$, and the best we could hope is that XZ is uniform on this interval. But this can be achieved by making sure that X only takes on the values $+1$ or -1.

PROBLEM 5.

(a)

$$\begin{aligned} I(X;Y) &= I(X_k, K; Y_k, K) = I(K; Y_k, K) + I(X_k; Y_K, K|K) = H(K) + I(X_k; Y_k|K) \\ &= h_2(\alpha) + \mathbb{P}_K[1].I(X_k; Y_k|K=1) + \mathbb{P}_K[2] I(X_k; Y_k|K=2) \\ &= h_2(\alpha) + \alpha.I(X_1; Y_1) + (1-\alpha) I(X_2; Y_2) \end{aligned}$$

(b) The distribution of X is determined by α and by the distributions of X_1 and X_2. It is clear from the expression in (a) that for any given α, $I(X;Y)$ is maximized when $I(X_1;Y_1)$ and $I(X_2;Y_2)$ are maximized, i.e., when the distribution of X_1 (resp. X_2) achieves the capacity of P_1 (resp. P_2). We conclude that the value of α in the capacity achieving distribution is the one that maximizes the function $f(\alpha) = h_2(\alpha) + \alpha C_1 + (1-\alpha)C_2$. The derivative of f is:

$$f'(\alpha) = -\log_2(\alpha) - \frac{1}{\ln 2} + \log_2(1-\alpha) + \frac{1}{\ln 2} + C_1 - C_2 = C_1 - C_2 + \log_2 \frac{1-\alpha}{\alpha}.$$

We have $f'(\alpha) = 0$ (resp. $f'(\alpha) > 0$, $f'(\alpha) < 0$) if $\alpha = \alpha^*$ (resp. $\alpha < \alpha^*$, $\alpha > \alpha^*$), where $\alpha^* = \dfrac{2^{C_1}}{2^{C_1} + 2^{C_2}}$. This means that $f(\alpha)$ is maximized at $\alpha = \alpha^*$. Therefore, the capacity achieving distribution is such that $\alpha = \dfrac{2^{C_1}}{2^{C_1} + 2^{C_2}}$ and X_1 (resp. X_2) achieves the capacity of the channel P_1 (resp. P_2).

(c) From (b), we have:

$$\begin{aligned} C &= -\frac{2^{C_1}}{2^{C_1}+2^{C_2}} \log_2 \frac{2^{C_1}}{2^{C_1}+2^{C_2}} - \frac{2^{C_2}}{2^{C_1}+2^{C_2}} \log_2 \frac{2^{C_2}}{2^{C_1}+2^{C_2}} + \frac{2^{C_1} C_1}{2^{C_1}+2^{C_2}} + \frac{2^{C_2} C_2}{2^{C_1}+2^{C_2}} \\ &= -\frac{2^{C_1}}{2^{C_1}+2^{C_2}} C_1 + \frac{2^{C_1}}{2^{C_1}+2^{C_2}} \log_2(2^{C_1}+2^{C_2}) - \frac{2^{C_2}}{2^{C_1}+2^{C_2}} C_2 \\ &\quad + \frac{2^{C_2}}{2^{C_1}+2^{C_2}} \log_2(2^{C_1}+2^{C_2}) + \frac{2^{C_1} C_1}{2^{C_1}+2^{C_2}} + \frac{2^{C_2} C_2}{2^{C_1}+2^{C_2}} \\ &= \log_2(2^{C_1}+2^{C_2}). \end{aligned}$$

Therefore, $2^C = 2^{C_1} + 2^{C_2}$.

PROBLEM 6. (a) Since X and Y are independent, $H(X,Y) = H(X) + H(Y)$. $H(X)$ can be found by the following steps.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{\infty} (1-p)^{i-1} p \log\left((1-p)^{i-1} p\right) \\ &= -\sum_{i=1}^{\infty} (1-p)^{i-1} p \left((i-1)\log(1-p) + \log p\right) \\ &= -p\log(1-p) \sum_{i=1}^{\infty} (1-p)^{i-1}(i-1) - p\log p \sum_{i=1}^{\infty} (1-p)^{i-1} \\ &= -(1-p)\log(1-p)/p - p\log p/p \\ &= h_2(p)/p. \end{aligned}$$

Similarly, $H(Y) = h_2(q)/q$, and $H(X,Y) = h_2(p)/p + h_2(q)/q$.

(b) Since $(X, Y) \to (2X + Y, X - 2Y)$ is a 1-to-1 transformation, $H(X, Y) = H(2X + Y, X - 2Y)$. To see this, we can write

$$\begin{aligned} H(X, Y, 2X + Y, X - 2Y) &= H(X, Y|2X + Y, X - 2Y) + H(2X + Y, X - 2Y) \\ &= H(2X + Y, X - 2Y|X, Y) + H(X, Y) \end{aligned}$$

As $H(X, Y|2X + Y, X - 2Y) = H(2X + Y, X - 2Y|X, Y) = 0$, we obtain $H(2X + Y, X - 2Y) = H(X, Y)$.

(c) Similar to part (a), $h(X, Y) = h(X) + h(Y)$. To find $h(X)$, consider the following steps.

$$\begin{aligned} h(X) &= -\int_0^\infty \lambda_X e^{-\lambda_X t} \log\left(\lambda_X e^{-\lambda_X t}\right) dt \\ &= -\int_0^\infty \lambda_X e^{-\lambda_X t} \log \lambda_X dt + \int_0^\infty \lambda_X^2 t e^{-\lambda_X t} dt \\ &= -\log \lambda_X - \lambda_X E[X] \\ &= 1 - \log \lambda_X \end{aligned}$$

as $E[X] = 1/\lambda_X$. Similarly, $h(Y) = 1 - \log \lambda_Y$ and $h(X, Y) = 2 - \log \lambda_X \lambda_Y$.

(d) Here, we cannot use the fact that $(X, Y) \to (2X + Y, X - 2Y)$ is a 1-to-1 transformation as $h(X, Y) \neq h(f(X, Y), g(X, Y))$ in general even if it is a 1-to-1 transformation. Instead, we use the fact that $h(\mathbf{Ax}) = h(\mathbf{x}) + \log |\mathbf{A}|$. Since we know that

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix},$$

we have $h(2X + Y, X - 2Y) = h(X, Y) + \log |\mathbf{A}| = 2 - \log \frac{\lambda_X \lambda_Y}{5}$.