ASSIGNMENT 4

Exercise 1 (Pure randomness from biased distributions)**.** Let $X_1, X_2, \ldots, X_n$ denote the outcomes of independent flips of a biased coin. Thus, for $i = 1, \ldots, n$ we have $\text{Pr}(X_i = 1) = p, \text{Pr}(X_i = 0) = 1 - p$, where p is unknown. We wish to obtain a sequence $Z_1, Z_2, \ldots, Z_K$ of fair coin flips from $X_1, X_2, \ldots, X_n$. To this end let $f : \mathcal{X}^n \to \{0,1\}^\star$ (where $\{0,1\}^\star = \{\Lambda, 0, 1, 00, 01, \ldots\}$ is the set of all finite length binary sequences including the null string Λ) be a mapping $f(X_1, X_2, \ldots, X_n) = (Z_1, Z_2, ..., Z_K)$, such that $Z_i \sim \text{Bernoulli}(1/2)$ and where K possibly depends on $(X_1, ..., X_n)$. For the sequence $Z_1, Z_2, ..., Z_K$ to correspond to fair coin flips, the map f from biased coin flips to fair flips must have the property that all 2^k sequences $(z_1, z_2, \ldots, z_k)$ of a given length k have equal probability (possibly 0). For example, for $n = 2$, the map $f(01) = 0, f(10) = 1, f(00) = f(11) = \Lambda$ has the property that $\text{Pr}(Z_1 = 1|K = 1) = \text{Pr}(Z_1 = 0|K = 1) = 1/2$.

a. Justify the following (in)equalities

$$\begin{aligned} nH_b(p) &\overset{(a)}{=} H(X_1, ..., X_n) \\ &\overset{(b)}{\geq} H(Z_1, Z_2, \ldots, Z_K, K) \\ &\overset{(c)}{=} H(K) + H(Z_1, Z_2, ..., Z_K|K) \\ &\overset{(d)}{=} H(K) + E(K) \\ &\overset{(e)}{\geq} E(K) \end{aligned}$$

where $E(K)$ denotes the expectation of K. Thus, on average, no more than $nH_b(p)$ fair coin tosses can be derived from $(X_1, ..., X_n)$.

b. Exhibit a good map f on sequences of length $n = 4$.

Solution. a. (a.) the X_i's are i.i.d. Bernoulli(p) distributed; (b) (Z^K, K) is a function of X^n; (c) chain rule; (d) given $K = k$, $(Z_1, Z_2, \ldots, Z_k)$ is an i.i.d. Bernoulli$(1/2)$ sequence, hence $H(Z_1, Z_2, ..., Z_K|K = k) = k$, from which the result follows; (d) non-negativity of the entropy.

b. One possibility is as follows. Let T_k be the set of binary sequences of length 4 with exactly k ones ($k \in \{0, 1, 2, \ldots, 4\}$). Observe that T_1 and T_3 each have four elements, and each contains equiprobable elements (obviously, the elements in T_1 have a different probability than those in T_3). We map the 4 elements in T_1 in 00, 01, 10, and 11, and similarly for T_3. I follows that, given $K = 2$, (Z_1, Z_2) are purely random. To see this note that for any pair of bit (i, j)

$$\begin{aligned} \text{Pr}((Z_1, Z_2) = (i,j)|K = 2) &= \text{Pr}((Z_1, Z_2) = (i,j)|X^4 \in T_1 \cup T_3) \\ &= \text{Pr}((Z_1, Z_2) = (i,j)|X^4 \in T_1)\text{Pr}(X^4 \in T_1|X^4 \in T_1 \cup T_3) \\ &+ \text{Pr}((Z_1, Z_2) = (i,j)|X^4 \in T_3)\text{Pr}(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}\text{Pr}(X^4 \in T_1|X^4 \in T_1 \cup T_3) + \frac{1}{4}\text{Pr}(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}. \end{aligned}$$

All the elements in T_0, T_2, and T_4 are mapped into Λ.

□

Exercise 2 (Capacity of two channels)**.** Consider two DMCs $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2, respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ are sent simultaneously, resulting in y_1, y_2. Find the capacity of this channel.

Solution. We have

$$p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2)$$

which implies the Markov chains

$$X_2 - X_1 - Y_1 \quad \text{and } X_1 - X_2 - Y_2.$$

Then,

$$\begin{aligned}
C &= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1, Y_2) \\
&= \max_{p(x_1,x_2)} I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&\overset{(a)}{=} \max_{p(x_1,x_2)} I(X_1; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&= \max_{p(x_1,x_2)} I(X_1; Y_1) + H(Y_2|Y_1) - H(Y_2|Y_1, X_1, X_2) \\
&\overset{(b)}{=} \max_{p(x_1,x_2)} I(X_1; Y_1) + H(Y_2|Y_1) - H(Y_2|X_2) \\
&\overset{(c)}{\leq} \max_{p(x_1,x_2)} I(X_1; Y_1) + H(Y_2) - H(Y_2|X_2) \\
&= \max_{p(x_1,x_2)} I(X_1; Y_1) + I(X_2; Y_2) \\
&\leq \max_{p(x_1,x_2)} I(X_1; Y_1) + \max_{p(x_1,x_2)} I(X_2; Y_2) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\
&= C_1 + C_2
\end{aligned}$$

where (a) is due to the Markov chain $X_2 - X_1 - Y_1$, (b) is due to the Markov chain $Y_2 - X_2 - (X_1, Y_1)$ and (c) is derived using conditioning inequality.

The equality can be achieved in all steps by choosing X_1 and X_2 independent, *i.e.*, $p(x_1, x_2) = p(x_1)p(x_2)$. So, $C = C_1 + C_2$. □

Exercise 3 (Choice of channels)**.** Find the capacity C of the union of two channels $(\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$, where at each time, one can send a symbol over channel 1 or channel 2 but not both. Assume that the output alphabets are distinct and do not intersect. Show that $2^C = 2^{C_1} + 2^{C_2}$. Thus, 2^C is the effective alphabet size of a channel with capacity C.

Solution. Let

$$\theta = \begin{cases} 1 & \text{with probability } p \\ 2 & \text{with probability } 1 - p \end{cases}$$

be the indicator that shows we are using which channel. Also, define $X = X_\theta$ and $Y = Y_\theta$. The capacity of the channel is thus

$$C = \max_{p(x)} I(X;Y).$$

Now notice that

$$I(X,\theta;Y) = I(\theta;Y) + I(X;Y|\theta) = I(X;Y) + I(\theta;Y|X).$$

So,

$$\begin{aligned}
I(X;Y) &= I(\theta;Y) + I(X;Y|\theta) - I(\theta;Y|X) \\
&\overset{(a)}{=} H(\theta) + I(X;Y|\theta) \\
&= H(\theta) + I(X_1;Y_1|\theta=1)Pr(\theta=1) + I(X_2;Y_2|\theta=2)Pr(\theta=2) \\
&= H(\theta) + I(X_1;Y_1)Pr(\theta=1) + I(X_2;Y_2)Pr(\theta=2) \\
&= H(\theta) + p\cdot I(X_1;Y_1) + (1-p)\cdot I(X_2;Y_2)
\end{aligned}$$

where (a) is due to the facts that $H(\theta|Y) = H(\theta|X) = 0$ since the outputs (and also inputs) are different for two channels.

Now, we have

$$\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p}[\max_{p(x_1,x_2)}[H(\theta) + p\cdot I(X_1;Y_1) + (1-p)\cdot I(X_2;Y_2)]] \\
&= \max_{p}[H(\theta) + p\cdot \max_{p(x_1)} I(X_1;Y_1) + (1-p)\cdot \max_{p(x_2)} I(X_2;Y_2)] \\
&= \max_{p}[H(\theta) + p\cdot C_1 + (1-p)\cdot C_2] \\
&= \max_{p}[-p\log p - (1-p)\log(1-p) + p\cdot(C_1 - C_2) + C_2]
\end{aligned}$$

Taking derivative with respect to p and let it to zero, we have

$$\log(\frac{p^*}{1-p^*}) = C_1 - C_2,$$

$$p^* = \frac{2^{C_1-C_2}}{2^{C_1-C_2}+1},$$

So,

$$\begin{aligned}
C &= -p^*\log(\frac{p^*}{1-p^*}) - \log(1-p^*) + p^*\cdot(C_1-C_2) + C_2 \\
&= -\log(1-p^*) + C_2
\end{aligned}$$

Hence,

$$2^C = 2^{C_2 - \log(1-p^*)} = \frac{2^{C_2}}{1-p^*} = 2^{C_2}\cdot(2^{C_1-C_2}+1) = 2^{C_1} + 2^{C_2}.$$

□

Exercise 4 (Z-channel)**.** The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, x, y \in \{0, 1\}$$

a. Find the capacity of the Z-channel and the maximizing input probability distribution.

b. Assume that we choose a $(2^{nR}, n)$ code at random, where each codeword is a sequence of fair coin tosses. This will not achieve capacity. Find the maximum rate R such that the probability of error $P_e^{(n)}$, averaged over the randomly generated codes, tends to zero as the block length n tends to infinity.

Solution. a. Let $p = Pr(X = 1)$. So, $Pr(Y = 1) = 1 - Pr(Y = 0) = \frac{p}{2}$

$$H(Y|X) = H(Y|X = 0)Pr(X = 0) + H(Y|X = 1)Pr(X = 1) = 0 + 1 \cdot p = p$$

$$H(Y) = h_b(\frac{p}{2})$$

$$I(X;Y) = H(Y) - H(Y|X) = h_b(\frac{p}{2}) - p$$

Taking derivative with respect to p, it can be seen that the mutual information gets maximized for $p = \frac{2}{5}$ and the capacity is thus $C \simeq 0.32$.

b. From the proof of channel coding theorem, it can be seen that if we choose the codewords with probability $p(x)$, the rate $I(X;Y)$ can be achieved. Here, we choose the codewords with $Pr(X = 0) = \frac{1}{2}$, so

$$I(X;Y) = H(Y) - H(Y|X) = h_b(\frac{1}{4}) - \frac{1}{2} = \frac{3}{2} - \frac{3}{4}\log 3$$

is achievable which is less than the capacity.

□

Exercise 5 (Unused symbols)**.** Show that the capacity of the channel with transition probability matrix

$$Q = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

is achieved by a distribution that places zero probability on one of input symbols. Give an intuitive reason as to why that symbol is not used. What is the capacity of this channel?

Solution. Let $P_X = (p_1, p_2, p_3)$ be a capacity-achieving distribution. Then,

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H\left(\frac{2}{3}p_1 + \frac{1}{3}p_2, \frac{1}{3}, \frac{1}{3}p_2 + \frac{2}{3}p_3\right) - (p_1 + p_3)H\left(\frac{2}{3}, \frac{1}{3}\right) - p_2 \log_2 3. \end{aligned}$$

Substituting $p_2 = 1 - (p_1 + p_3)$, we obtain

$$I(X;Y) = H\left(\frac{1}{3} + \frac{1}{3}(p_1 - p_3), \frac{1}{3}, \frac{1}{3} + \frac{1}{3}(p_1 - p_3)\right) - (p_1 + p_3)H\left(\frac{2}{3}, \frac{1}{3}\right) - (1 - (p_1 + p_3)) \log_2 3.$$

For a fixed value of $p_1 + p_3$, the above expression is maximized when $p_1 = p_3$. Then, we have

$$\begin{aligned} I(X;Y) &= \log_2 3 - (p_1 + p_3)H\left(\frac{2}{3}, \frac{1}{3}\right) - (1 - (p_1 + p_3)) \log_2 3 \\ &= (p_1 + p_3)\left\{\log_2 3 - H\left(\frac{2}{3}, \frac{1}{3}\right)\right\}. \end{aligned}$$

Since the second term in the above product is positive, $I(X;Y)$ is maximized when $p_1 + p_3$ is maximized. Thus, set $p_1 + p_3 = 1$ which implies that $P_X = (1/2, 0, 1/2)$. This yields the capacity to be

$$C = \log_2 3 - H\left(\frac{2}{3}, \frac{1}{3}\right) = \frac{2}{3} \text{ bits.}$$

□

Exercise 6 (Erasures and errors in a binary channel)**.** A binary erasure channel with erasure probability β, denoted BEC(β) has output alphabet $\{0, 1, e\}$ and transitions given by $P(0|0) = P(1|1) = 1 - \beta$, and $P(e|0) = P(e|1) = \beta$ where e is the erasure symbol.

a. Show that the capacity of BEC(β) is $1 - \beta$.

b. Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be α and the probability of erasure be β, which means that when we send symbol 0, with probability $1 - \alpha - \beta$ we receive symbol 0, with probability α we receive symbol 1 and with probability β we receive an erasure symbol. Find the capacity of this channel.

Solution. a. We have

$$\begin{aligned} P(Y = e) &= P(Y = e, X = 0) + P(Y = e, X = 1) \\ &= P(X = 0)\beta + P(X = 1)\beta \\ &= \beta. \end{aligned}$$

Since

$$\begin{aligned} P(X = 0|Y = e) &= \frac{P(Y = e|X = 0)P(X = 0)}{P(Y = e)} \\ &= \frac{\beta P(X = 0)}{\beta} \\ &= P(X = 0), \end{aligned}$$

we have $H(X|Y = e) = H(X)$. Observe that

$$H(X|Y = 0) = H(X|Y = 1) = 0.$$

Hence,

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= (1 - \beta)H(X), \end{aligned}$$

which is maximized when $X \sim Ber(1/2)$ yielding the result.

b. The alphabet of $\mathcal{Y} = \{0, e, 1\}$ and the transition probability matrix is

$$Q(y|x) = \begin{pmatrix} 1-\alpha-\epsilon & \alpha & \epsilon \\ \epsilon & \alpha & 1-\alpha-\epsilon \end{pmatrix}$$

Due to the symmetry of transition probability matrix for inputs 0 and 1, $I(X;Y)$ is the same for $Pr(X=0) = p$ and $Pr(X=1) = 1-p$, so is symmetric with respect to $\frac{1}{2}$, moreover $I(X;Y)$ is concave with respect to p. These, yields that $I(X;Y)$ is maximized for $p = \frac{1}{2}$. With $p = \frac{1}{2}$, we have

$$Pr(Y=0) = \frac{1}{2}(1-\alpha)$$

$$Pr(Y=1) = \frac{1}{2}(1-\alpha)$$

$$Pr(Y=e) = \alpha$$

$$H(Y) = -(1-\alpha)\log(\frac{1-\alpha}{2}) - \alpha\log(\alpha)$$

$$H(Y|X) = H(1-\alpha-\epsilon, \epsilon, \alpha)$$

$$C = H(Y) - H(Y|X)$$

□

Exercise 7 (Binary multiplier channel)**.** Consider the channel $Y = X \cdot Z$, where X and Z are independent binary random variables and $Z \sim Ber(\alpha)$. [i.e., $P(Z=1) = \alpha$].

a. Find the capacity of this channel and the maximizing distribution on X.

b. Now suppose that the receiver can observe Z as well as Y. What is the capacity?

Solution. a. The transition probability matrix is

$$Q = \begin{pmatrix} 1 & 0 \\ 1-\alpha & \alpha \end{pmatrix}$$

So, if $Pr(X=1) = p$, then

$$H(Y|X) = H(Y|X=0)Pr(X=0) + H(Y|X=1)Pr(X=1) = 0\cdot(1-p) + h_b(\alpha)\cdot p = ph_b(\alpha).$$

$$H(Y) = h_b(\alpha \cdot p)$$

$$I(X;Y) = H(Y) - H(Y|X) = h_b(\alpha \cdot p) - ph_b(\alpha)$$

Taking derivative with respect to p and let it equal to zero, we have

$$\alpha \log\left(\frac{1-\alpha p^*}{\alpha p^*}\right) = h_b(\alpha),$$

$$p^* = \frac{1}{\alpha\left(2^{\frac{h_b(\alpha)}{\alpha}} + 1\right)}.$$

and

$$\begin{aligned} C = I(X;Y)|_{p=p^*} &= h_b(\alpha \cdot p^*) - p^* h_b(\alpha) \\ &= p^* \alpha \log\left(\frac{1-\alpha p^*}{\alpha p^*}\right) - \log(1-\alpha p^*) - p^* h_b(\alpha) \\ &= -\log(1-\alpha p^*) \\ &= \log\left(\frac{2^{\frac{h_b(\alpha)}{\alpha}}+1}{2^{\frac{h_b(\alpha)}{\alpha}}}\right) \end{aligned}$$

Note that for $\alpha = \frac{1}{2}$, this channel is the same as Z-channel.

b. In this case,

$$C = \max_{p(x)} I(X;Y,Z) = \max_{p(x)} [I(X;Z) + I(X;Y|Z)] = \max_{p(x)} I(X;Y|Z).$$

Assuming $Pr(X=1) = p$,

$H(Y|Z) = H(Y|Z=0)Pr(Z=0) + H(Y|Z=1)Pr(Z=1) = 0 \cdot (1-\alpha) + H(X|Z=1) \cdot \alpha = \alpha H(X) = \alpha h_b(p)$

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

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) = \alpha h_b(p)$$

which is maximized for $p = \frac{1}{2}$ and hence, $C = \alpha$.

□

Exercise 8 (Memoryless channels without feedback)**.** Consider a channel W given by $(\mathcal{X}^n, P(y^n|x^n), \mathcal{Y}^n)$.

1. The channel is said to be memoryless if
$$P(y_i|x^i, y^{i-1}) = P(y_i|x_i) \quad 1 \le i \le n.$$
2. The channel is said to be used *without feedback* if
$$P(x_i|x^{i-1}, y^{i-1}) = P(x_i|x^{i-1}) \quad 1 \le i \le n.$$

Show that if the channel is memoryless and used without feedback then

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x_i).$$

Hint – Expand $P(x^n, y^n)$ using Bayes' rule.

Solution.

$$\begin{aligned} P(x^n, y^n) &= \prod_{i=1}^{n} P(x_i, y_i|x^{i-1}, y^{i-1}) \\ &= \prod_{i=1}^{n} P(x_i|x^{i-1}, y^{i-1}) P(y_i|x^i, y^{i-1}) \\ &= \prod_{i=1}^{n} P(x_i|x^{i-1}) P(y_i|x_i) \\ &= P(x^n) \prod_{i=1}^{n} P(y_i|x_i), \end{aligned}$$

whereby the result follows. We have used conditions 1 and 2 in the step before the last. □

Exercise 9. (Conditional KL divergence) The conditional KL divergence between P and Q, probability distributions on $(\mathcal{X} \times \mathcal{Y})$, is given by

$$D(P_{X|Y}||Q_{X|Y}|Y) = \sum_{x,y} P(x,y) \log \frac{P(x|y)}{Q(x|y)}.$$

a. Prove that like the KL divergence, the conditional KL divergence is also non-negative.

b. Prove that

$$I(X;Y) = \max_{V(x|y)} \sum_{x,y} Q(x)P(y|x) \log \frac{V(x|y)}{Q(x)},$$

and the maximum is attained by $V^*(x|y) = \frac{Q(x)P(y|x)}{\sum_x Q(x)P(y|x)}$. Here, the mutual information $I(X;Y)$ is calculated with respect to the joint distribution $Q(x)P(y|x)$.

Solution. a. $D(P_{X|Y}||Q_{X|Y}|Y) = \sum_{x,y} P(x,y) \log \frac{P(x|y)}{Q(x|y)} = \mathbb{E}_y[D(P_{X|Y=y}||Q_{X|Y=y})] \geq 0.$

a. By the log-sum inequality, we have

$$\begin{aligned} D(P_{X|Y}||Q_{X|Y}|Y) &= \sum_{x,y} P(x,y) \log \frac{P(x|y)}{Q(x|y)} \\ &= \sum_y P(y) \sum_x P(x|y) \log \frac{P(x|y)}{Q(x|y)} \\ &\geq \sum_y P(y) \left\{ \left(\sum_x P(x|y) \right) \log \frac{\sum_x P(x|y)}{\sum_x Q(x|y)} \right\} \\ &\geq 0. \end{aligned}$$

b. We have

$$\begin{aligned} \sum_{x,y} Q(x)P(y|x) \log \frac{V(x|y)}{Q(x)} &= \sum_{x,y} Q(x)P(y|x) \log \left(\frac{P(y|x)}{\sum_x Q(x)P(y|x)} \frac{V(x|y) \sum_x Q(x)P(y|x)}{Q(x)P(y|x)} \right) \\ &= I(X;Y) - \sum_{x,y} Q(x)P(y|x) \log \frac{V^*(x|y)}{V(x|y)} \\ &= I(X;Y) - D(V^*||V|Y). \end{aligned}$$

Since $D(V^*||V|Y) \geq 0$, we have

$$\sum_{x,y} Q(x)P(y|x) \log \frac{V(x|y)}{Q(x)} \leq I(X;Y)$$

and equality is attained by setting $V(\cdot|\cdot) = V^*(\cdot|\cdot)$.

□

Exercise 10. (Capacity-achieving output distribution is unique) Consider a discrete memoryless channel and let $P_X^{(1)}$ and $P_X^{(2)}$ by two capacity-achieving input distributions. Define

$$P_X = \theta P_X^{(1)} + (1-\theta) P_X^{(2)}.$$

a. Let $X^{(1)} \sim P_X^{(1)}$ and $X^{(2)} \sim P_X^{(2)}$. Define

$$X = ZX^{(1)} + (1-Z)X^{(2)}$$

where $Z \sim Ber(\theta)$. Convince yourself that $X \sim P_X$.

b. Does $Z - X - Y$ form a Markov chain? Namely, verify that for all x, y, z,

$$P(z, y|x) = P(z|x)P(y|x).$$

c. Show that $I(Z;Y) = 0$. *Hint* – Apply chain rule to $I(X, Z; Y)$

d. Conclude that the capacity-achieving output distribution is unique.

Solution. Let $X \sim Q$ and $Z \sim P_Z$. We have

$$Q = \sum_z Q_{X|Z=z} P_Z(z).$$

Since $X = X^{(1)}$ when $Z = 1$ and $X = X^{(2)}$ when $Z = 0$, we have

$$Q_{X|Z=1} = P_X^{(1)} \text{ and } Q_{X|Z=0} = P_X^{(2)},$$

whereby $Q = P_Z(1)P_X^{(1)} + P_Z(0)P_X^{(2)} = P_X$. Given $X = X^{(1)}$, we have $Z = 1$ and $X = X^{(2)}$, we have $Z = 0$. Therefore, $H(Y|X, Z) = H(Y|X)$ and hence we have the Markov chain $Z - X - Y$. Now, we expand $I(X, Z; Y)$ in two different ways. We have

$$I(X, Z; Y) = I(X; Y) + I(Z; Y|X)$$

and

$$I(X, Z; Y) = I(Z; Y) + I(X; Y|Z).$$

By the Markov chain above, $I(Z; Y|X) = 0$. Given $Z = 0$, $I(X; Y|Z = 0)$ is the mutual information between X and Y where $X \sim P_X^{(1)}$, which is equal to the capacity C. The same argument holds for $Z = 1$. Also, P_X which is a convex combination of two capacity-achieving distributions, is a capacity-achieving distribution. Therefore,

$$I(Z; Y) = I(X; Y) - I(X; Y|Z) = C - C = 0.$$

Thus, the distribution of Y is independent of Z and hence unique. □