Throughout the homework we assume that alphabets are countable.

1. **Reproving Fano's Inequality.**
Consider the Markov chain $X \to Y \to \hat{X}$, where X takes values in $\mathcal{X}$ with $|\mathcal{X}| < \infty$, and define the probability of error $P_e := \mathbb{P}(\hat{X} \neq X)$.

(a) Define the binary random variable $E := \mathbb{1}_{\{\hat{X} \neq X\}}$ and show that $\mathbb{E}[E] = P_e$ and $\texttt{var}(E) = P_e(1 - P_e)$.

Solution.
We can view E as a Bernoulli random variable whose probability of success is the probability of the event the indicator function measures. That is, $E = \texttt{Ber}(\mathbb{P}(\hat{X} = X)) = \texttt{Ber}(P_e)$. It then immediately follows that $\mathbb{E}[E] = P_e$ and $\texttt{var}(E) = P_e(1 - P_e)$ as desired. ∎

(b) Use the chain rule of entropy to show $H(X|\hat{X}) + H(E|X, \hat{X}) = H(E|\hat{X}) + H(X|E, \hat{X})$.

Solution.
Begin by noting that

$$H(E, X|\hat{X}) = H(X|\hat{X}) + H(E|X, \hat{X}),$$

and

$$H(E, X|\hat{X}) = H(E|\hat{X}) + H(X|E, \hat{X})$$

giving that

$$H(X|\hat{X}) + H(E|X, \hat{X}) = H(E|\hat{X}) + H(X|E, \hat{X})$$

∎

(c) Prove the following: $H(X|E, \hat{X}) \leq P_e \log|\mathcal{X}|$.

Solution.

$$\begin{aligned}
H(X|E, \hat{X}) &= \mathbb{P}(\hat{X} = X)H(X|E = 0, \hat{X}) + \mathbb{P}(\hat{X} \neq X)H(X|E = 1, \hat{X}) \\
&= 0 + \mathbb{P}(\hat{X} \neq X)H(X|E = 1, \hat{X}) \\
&\leq \mathbb{P}(\hat{X} \neq X) \log|\mathcal{X}| \\
&= P_e \log|\mathcal{X}|
\end{aligned}$$

∎

(d) Prove $H_b(P_e) + P_e \log|\mathcal{X}| \geq H(X|\hat{X})$, where H_b is the binary entropy function.

Solution.
Since E is a deterministic function of $(X, \hat{X})$, we have that $H(E|X, \hat{X}) = 0$. Thus

$$\begin{aligned} H(X|\hat{X}) &= H(E|\hat{X}) + H(X|E, \hat{X}) \\ &\leq H(E) + P_e \log|\mathcal{X}| \\ &\leq H_b(P_e) + P_e \log|\mathcal{X}| \end{aligned}$$

■

(e) Using the data processing inequality and the results above, show that $P_e \geq \frac{H(X|Y)-1}{\log|\mathcal{X}|}$

Solution.
We note that $X \leftrightarrow Y \leftrightarrow \hat{X}$ is a Markov chain giving that

$$I(X;Y) \geq I(X;\hat{X}) \iff H(X|Y) \leq H(X|\hat{X})$$

via the data-processing inequality. Thus

$$H(X|Y) \leq H(X|\hat{X}) \leq H_b(P_e) + P_e \log|\mathcal{X}|$$

$$\implies P_e \geq \frac{H(X|Y) - H_b(P_e)}{log|X|} \geq \frac{H(X|Y) - 1}{log|X|}$$

■

2. **General Fano's Inequality.**
Let $X, Y \sim P_{X,Y} \sim \mathcal{P}(\mathcal{X} \times \mathcal{X})$. Define $Q_{X,Y} = P_X \otimes P_Y$ and let $p := P_{X,Y}(X = Y)$ and $q := Q_{X,Y}(X = Y)$. In words, p and q are the probabilities that $X = Y$ under the joint law $P_{X,Y}$ or the product of marginals law $Q_{X,Y}$, respectively.

(a) Prove that $I(X;Y) \geq p \log\left(\frac{1}{q}\right) - H_b(p)$, where H_b is the binary entropy function.

Solution.
We have $I(X;Y) = D_{\mathsf{KL}}\left(P_{XY} \| Q_{XY}\right)$. We use f-divergence DPI with the mapping that takes $x = y$ and get

$$\begin{aligned} I(X;Y) &= D_{\mathsf{KL}}\left(P_{XY} \| Q_{XY}\right) \\ &\geq D_{\mathsf{KL}}\left(Ber(p) \| Ber(q)\right) \\ &= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \\ &= p \log \frac{1}{q} - (p \log \frac{1}{p} + (1-p) \log \frac{1}{(1-p)}) + (1-p) \log \frac{1}{(1-q)} \\ &= p \log \frac{1}{q} - H_b(p) + (1-p) \log \frac{1}{(1-q)} \\ &\geq p \log \frac{1}{q} - H_b(p) \end{aligned}$$

where the last inequality holds because $\log \frac{1}{(1-q)}$ is non-negative, so we get the desired result $I(X;Y) \geq p \log \frac{1}{q} - H_b(p)$ ■

(b) Assume $P_X = \texttt{Unif}(\mathcal{X})$ (clearly $|\mathcal{X}| < \infty$) and obtain the regular Fano's inequality from part (a), i.e.,

$$H(X|Y) \leq P_{X,Y}(X \neq Y) \log|\mathcal{X}| + H_b(P_{X,Y}(X \neq Y)).$$

Solution.
First, note that since P is uniform, then $q = Q_{X,Y}(X = Y) = \frac{1}{|\mathcal{X}|}$. From part a, we have $I(X;Y) \geq p \log \frac{1}{q} - H_b(p)$. Using $I(X;Y) = H(X) - H(X|Y)$, we have

$$\begin{aligned}
H(X|Y) &\leq H(X) - p \log \frac{1}{q} + H_b(p) \\
&\leq \log |\mathcal{X}| - p \log |\mathcal{X}| + H_b(p) \\
&= (1-p) \log |\mathcal{X}| + H_b(p) \\
&= Pxy(X \neq Y) \log|\mathcal{X}| + H_b(p) \\
&= Pxy(X \neq Y) \log|\mathcal{X}| H_b(Pxy(X \neq Y))
\end{aligned}$$

where the last equality comes from the symmetry of H_b ($H_b(\alpha) = H_b(1-\alpha). \forall \alpha \in (0,1)$) ■

3. **Application of Fano's Inequality.**
Let $\mathcal{X} = \{1, 2, ..., m\}$ and $X \sim P$ with PMF $p(i) = p_i$ for all $i \in \mathcal{X}$ such that $p_1 \geq p_2 \geq \ldots \geq p_m$.

(a) Consider deterministic predictors of the form $\hat{X} = i$. Find the predictor with minimum probability of error P_e.

Solution.
The predictor with minimum probability of error is $\hat{X} = 1$. I.e. $\hat{P}_e = 1 - p_1$. This is obvious since p_1 has the highest probability. ■

(b) Maximize $H(P)$ subject to the constraint that the probability of error of the estimator found in part (a) remains unchanged.

Solution.
Since we have to preserve the probability of error, then we need to keep the probability of the message being 1 to be the same p_1. We construct a new distribution P that maximizes $H(P)$ while maintaining the error probability by defining the new pmf $(\hat{p_1}, ..., \hat{p_m})$ as follows:

$$\hat{p_1} = p_1, \quad \text{and} \quad \hat{p_i} = \frac{1 - p_i}{m - 1} \quad \forall i \neq 1$$

. This distribution maximizes the entropy since the probability of the messages 2,...,m is uniform while keeping the probability of 1 - and hence the probability of error - fixed. We get

$$\begin{aligned}
H(P) &= \hat{p}_1 \log \frac{1}{\hat{p}_1} + (\hat{p}_1) \sum_{i=2}^{m} \hat{p}_i (\log \frac{1}{\hat{p}_i} + \log \frac{1}{1-\hat{p}_1}) \\
&= \hat{p}_1 \log \frac{1}{\hat{p}_1} + (\hat{p}_1) \sum_{i=2}^{m} \hat{p}_i (\log(m-1) + \log \frac{1}{1-\hat{p}_1}). \\
&= H_b(p_1) + \log(m-1)
\end{aligned}$$

■

(c) Use part (b) to derive a bound on P_e in terms of entropy.

Solution.

From part (b) we can conclude that

$$H(X) \leq H(P) = H(\hat{P}_e) + \hat{P}_e \log_2 |m-1|$$

$$\implies \hat{P}_e \geq \frac{H(X) - H(\hat{P}_e)}{\log_2 |m-1|}$$

$$\geq \frac{H(X) - 1}{\log_2 |m-1|}$$

■

4. **Alphabet of Noise.**
Let $\mathcal{X} = \{0, 1, 2, 3, 4\}$. Consider the channel $Y = X + Z$ where Z is uniformly distributed over $\mathcal{Z} = \{z_1, z_2, z_3\}$. Here $z_1, z_2, z_3 \in \mathbb{Z}$.

(a) What is the maximum capacity of this channel? Provide a distribution on $\mathcal{X}$ and a choice of alphabet $\mathcal{Z}$ that achieves that capacity.

Solution. Note that $I(X;Y) \leq \min\{H(X), H(Y)\}$, and since Y is the sum of independent random variables, we have $H(Y) \geq \max\{H(X), H(Z)\}$, so we have $I(X;Y) \leq H(X) \leq \log |\mathcal{X}| = \log 5$. By choosing $\mathcal{Z} = \{-5, 0, 5\}$ and $P_X = Unif(X)$, then we get $Y = Unif(\{-5, -4, ..., 0, 1, ..., 9\})$, so we have $H(Y) = \log 15$, and we have $H(Y|X) = H(Z) = \log 3$ (as shown on HW3). So we have

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|Z) \\
&= \log 15 - \log 3 \\
&= \log 5
\end{aligned}$$

which is the upper bound that we shown. So the maximum capacity is $\log 5$, with maximizing X distribution $Unif(\mathcal{X})$ and $\mathcal{Z} = \{-5, 0, 5\}$ ■

(b) What is the minimum capacity of this channel? Provide a distribution on $\mathcal{X}$ and a choice of alphabet $\mathcal{Z}$ that achieves that capacity.

Solution. To minimize the channel capacity, we choose $\mathcal{Z}$ s.t. intuitively it increase the interference between the messages and at the same time keep the number of possible values of Y to a minimum. So we choose $\mathcal{Z} = \{-1, 0, 1\}$. With this choice of $\mathcal{Z}$, we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(Z) \\ &\geq \log(7) - \log(3) \approx 1.222 \end{aligned}$$

since Y can have at most 7 values and it's entropy is bounded with $\log 7$. However, making Y to be uniform is not achievable with any choice of P_X, so that upper bound is not attainable. So to optimize the mutual information, we reformulate it as a classical optimization problem. We want P_Y to be as close to uniform distribution as possible, so we construct the pmf of X as $p_0 = p_4 = \frac{y}{2y+x}$ and $p_2 = \frac{x}{2y+x}$ with $p_1 = p_3 = 0$. And since the only the ratio of x and y is what matters, we fix $y = 1$ and maximize the mutual information with the expression

$$4\left(\tfrac{y}{6y+3x}\log_2\left(\tfrac{6y+3x}{y}\right)\right) + \tfrac{x}{6y+3x}\log_2\left(\tfrac{6y+3x}{x}\right) + 2\left(\tfrac{x+y}{6y+3x}\log_2\left(\tfrac{6y+3x}{x+y}\right)\right) - \log_2(3)$$

With the constraint $x \geq 0$. We obtain a maximum at $x = 0.618, y = 1$ with $I(X;Y) \approx 1.157$ ■

5. **Union of Channels.**
In this problem we want to find a capacity C of the union of two channels. Specifically, let $\mathcal{X}_1 = \{1, ..., m\}$ and $\mathcal{X}_2 = \{m+1, ..., n\}$ and channels $(\mathcal{X}_1, p_{Y_1|X_1}, \mathcal{Y}_1)$, $(\mathcal{X}_2, p_{Y_2|X_2}, \mathcal{Y}_2)$, where at each time, one can send a symbol over channel 1 or channel 2 but not both.

(a) Let $X_1 \sim P_1 \in \mathcal{P}(\mathcal{X}_1)$, $X_2 \sim P_2 \in \mathcal{P}(\mathcal{X}_2)$, and

$$X = \begin{cases} X_1, & \text{with probability } p, \\ X_2, & \text{with probability } 1-p. \end{cases}$$

Find $H(X)$ in terms of $H(X_1), H(X_2)$, and p.

Solution.
Define an indicator of sorts, c, so that

$$c = \begin{cases} 1, & X = X_1 \\ 2, & X = X_2 \end{cases}.$$

Then

$$\begin{aligned}
H(X) = H(X, c) &= H(c) + H(X|c) \\
&= H(c) + p \cdot H(X|c = 1) + (1 - p) \cdot H(X|c = 2) \\
&= H_b(p) + p \cdot H(X_1) + (1 - p) \cdot H(X_2)
\end{aligned}$$

■

(b) Maximize over p to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$. One can view $2^{H(X)}$ as the effective alphabet size.

Solution.
After taking the derivative with respect to p and evaluating it at zero we obtain a maximum at $p = \hat{p}$, where

$$\hat{p} = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}.$$

Plugging $\hat{p}$ back into the expression for $H(X)$ we see that

$$\begin{aligned}
H(X) &= H_b(\hat{p}) + \hat{p} \cdot H(X_1) + (1 - \hat{p}) \cdot H(X_2) \\
&= \log_2 \left(2^{H(X_1)} + 2^{H(X_2)}\right),
\end{aligned}$$

giving that

$$2^{H(X)} = 2^{H(X_1)} + 2^{H(X_2)}.$$

Since this was obtained at the max value of $\hat{p}$, we note that equality only holds in that case. That is, for a general $p \neq \hat{p}$ we have

$$2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}.$$

■

(c) Find the capacity of the union of two channels C.

Solution.
Begin by finding the mutual information of X and Y.

$$\begin{aligned}
I(X; Y) &= H(Y) - H(Y|X) \\
&= H(Y, c) - H(Y|X, c) \\
&= H(c) + p \cdot H(Y|c = 1) + (1 - p) \cdot H(Y|c = 2) \\
&\qquad - p \cdot H(Y|X, c = 1) - (1 - p) \cdot H(Y|X, c = 2) \\
&= H_b(p) + p \cdot \big(H(Y_1) - H(Y_1|X_1)\big) + (1 - p) \cdot \big(H(Y_2) - H(Y_2|X_2)\big) \\
&= H_b(p) + p \cdot I(X_1; Y_1) + (1 - p) \cdot I(X_2; Y_2).
\end{aligned}$$

We now note that to obtain capacity of the channel we would have to have maximum capacity on our other two channels. Thus the optimization is left with respect to p and is of the form

$$C = \max_p \left\{ H_b(p) + p \cdot C_1 + (1-p) \cdot C_2 \right\}.$$

This is an identical optimization as above and we thus obtain $p = \hat{p}$ as our maximizing p-value, where

$$\hat{p} = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}.$$

Plugging $\hat{p}$ into our capacity expression yields

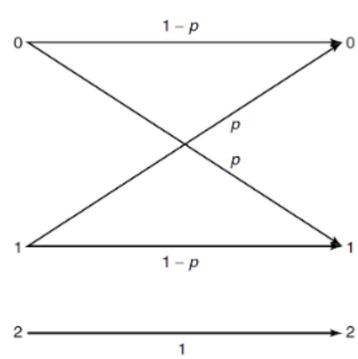
$$C = \log_2 \left(2^{C_1} + 2^{C_2} \right)$$

giving that

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

which makes sense! If we view 2^C as the effective alphabet size of a channel with capacity C then at each point in time, since we can only transmit through one channel at a time, we have an effective alphabet size of $2^{C_1} + 2^{C_2}$ thus leading to $2^C = 2^{C_1} + 2^{C_2}$. ■

(d) Use part (c) to compute the capacity of the following channel.



Solution.
We first note that the second channel has a capacity of zero, i.e. $C_2 = 0$. Using our result from (c) and the fact that a binary symmetric channel has max capacity $C_1 = 1 - H_b(p)$ we obtain

$$C = \log_2 \left(2^{1-H_b(p)} + 2^0 \right).$$

■

6. **Preprocessing the Output.**
Consider a discrete memoryless channel (DMC) $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$.

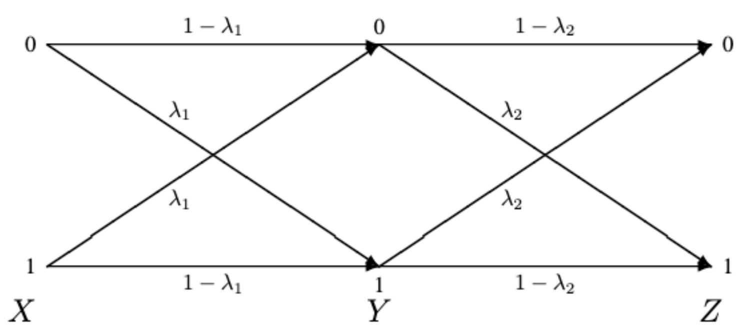
(a) Prove that it is impossible to strictly increase the channel capacity $\max_{P_X} I(X;Y)$ by preprocessing the output Y by forming $\hat{Y} = g(Y)$, giving rise to a new (effective) channel $P_{\hat{Y}|X}$.

Solution.
Note that the capacity $C = \sup_{P_X} I(X,Y)$. However, it follows from DPI of mutual information that $I(X,Y) \geq I(X, g(Y))$. Denote the new capacity with C'. We have $C' = \sup_{P_X} I(X, g(Y)) \leq \sup_{P_X} I(X,Y) = C$. ■

(b) Under what conditions does preprocessing not strictly decrease the capacity?

Solution.
The preprocessing will not strictly decrease the capacity if DPI holds with equality, and we have DPI hold with equality if and only if the processing function g is a bijection (since we will not lose any information then). ■

7. **Cascaded BSCs.**
Consider two DMCs $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ and $(\mathcal{Y}, \mathcal{Z}, P_{Z|Y})$. Let $P_{Y|X}$ and $P_{Z|Y}$ be binary symmetric channels with crossover probabilities λ_1 and λ_2 respectively.



(a) What is the capacity C_1 of $P_{Y|X}$?

Solution.
$P_{Y|X}$ is a binary symmetric channel, so its capacity is $C_1 = 1 - H_b(\lambda_1)$ as derived in lecture. It follows since $I(X;Y) = H(Y) - H(Y|X)$ where $H(Y|X) = H_b(\lambda_1)$ and $H(Y)$ for a binary symmetric channel is always maximized at 1 when choosing X to be uniform. ■

(b) What is the capacity C_2 of $P_{Z|Y}$?

Solution.
Using the same argument as part a, we have $C_2 = 1 - H_b(\lambda_2)$ ■

(c) We now cascade these channels, obtaining a new effective channel $Q_{Z|X}$ given by

$$Q_{Z|X}(z|x) = \sum_y P_{Y|X}(y|x)P_{Z|Y}(z|y), \quad \forall x, z \in \{0, 1\}.$$

What is the capacity C_3 of $Q_{Z|X}$? Show that $C_3 \leq \min\{C_1, C_2\}$.

Solution.
We have $C_3 = \sup_{P_X} I(Z; X) = \sup_{P_X} H(Z) - H(Z|X)$. First, note that $H(Z|X)$ is independent of P_X and we will compute it later. Now, note that $H(Z) \leq \log 2 = 1$. We can achieve that entropy by choosing $P_X = Ber(\frac{1}{2})$, since for a BSC $P_{Y|X}$, setting the input to be $Ber(\frac{1}{2})$ makes the output Y to be also $Ber(\frac{1}{2})$. So inductively, we also get $Z = Ber(\frac{1}{2})$ which achieves the upper bound of 1.
Now we consider $H(Z|X)$. We have $H(Z|X) = \sum_{x\in\{0,1\}} P(X = x)H(Z|X = x)$, but because of the symmetry of the channel and that entropy of binary alphabets is also symmetric, we get $H(Z|X) = H(Z|X = 0)$. Computing $P(Z|X = 0)$ we have

$$P(Z = 0|X = 0) = (1 - \lambda_1)(1 - \lambda_2) + \lambda_1\lambda_2$$

So $H(Z|X) = H_b((1 - \lambda_1)(1 - \lambda_2) + \lambda_1\lambda_2)$ and we have

$$C_3 = 1 - H_b((1 - \lambda_1)(1 - \lambda_2) + \lambda_1\lambda_2)$$

Now we relate C_3 to C_1 and C_2. First, define $C(x) = 1 - H_b(x)$ for $x \in (0, 1)$. Note that C is concave (since H_b is convex). So we have

$$\begin{aligned} C(\alpha x + (1 - \alpha)x) &\leq \alpha C(x) + (1 - \alpha)C(1 - x) \\ &= \alpha C(x) + (1 - \alpha)C(x) && \text{from the symmetry of } C \\ &= C(x) \end{aligned}$$

and by setting $\alpha = \lambda_1$ and $x = \lambda_2$, we have

$$C_3 = C((1 - \lambda_1)(1 - \lambda_2) + \lambda_1\lambda_2) \leq C(\lambda_2) = C_2$$

and similarly, by setting $\alpha = \lambda_2$ and $x = \lambda_1$ we have $C_3 \leq C_1$, so this gives us $C_3 \leq \min\{C_1, C_2\}$ as desired. ∎

(d) Now let us actively intervene between channels 1 and 2, rather than passively forwarding the output Y of $P_{Y|X}$ through $P_{Z|Y}$. What is the capacity of channel 1 followed by channel 2 if you are allowed to decode the output Y^n of channel 1 and then re-encode it as $\tilde{Y}^n$ for transition over channel 2?

Solution.
Consider the re-encoding scheme $\tilde{Y} = g(Y)$. For any mapping g on binary alphabet, we can represent it with a binary symmetric channel without a loss of generality, where the flipping parameter $\alpha = ¶(g(Y) \neq Y)$. As a result, introducing such re-encoding is equivalent to

adding a third additional channel between the two channels we have in part c. So now we essentially have three cascaded BSCs. To compute the capacity of this channel, note that a cascade of BSCs is a BSC, so we can proceed by computing the equivalent BSC to the first pair of BSCs, and then compute the capacity of the cascade of that equivalent BSC and the third BSC.

We will avoid explicitly computing the value for readability but we will state the steps. The capacity of the first channel cascaded with the re-encoding BSC $\hat{C}_1 = 1 - H_b(\mu_1)$ where $\mu_1 = (\lambda_1\alpha + (1-\lambda_1)(1-\alpha)$. Now, the capacity of the cascade of three channels end up being $\hat{C}_2 = 1 - H_b(\mu_2)$ where $\mu 2 = \mu_1\lambda_2 + (1-\mu_1)(1-\lambda_2)$ which is the capacity we wanted to compute.

Finally, note that the new capacity $\hat{C}_2 \leq C_3$ (where C_3 is the capacity form part c) and it holds with equality iff the encoding function is deterministic and bijective. ■

(e) What is the capacity of the cascade in part (c) if the end receiver can view *both* Y and Z?

Solution. The capacity will be C_1. This is because X is independent from Z given Y, giving us

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

which correspond to the mutual information of the first channel. ■

8. **Product Channel.**
Consider two DMCs $(\mathcal{X}_1, \mathcal{Y}_1, P_{Y_1|X_1})$ and $(\mathcal{X}_2, \mathcal{Y}_2, P_{Y_2|X_2})$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2, P_{Y_1,Y_2,|X_1,X_2})$, where $P_{Y_1,Y_2,|X_1,X_2} = P_{Y_1|X_1}(y_1|x_1)P_{Y_2|X_2}(y_2|x_2)$, for all (x_1, x_2, y_1, y_2), is formed. Find the capacity of this channel in terms of C_1 and C_2.

Solution.
Begin by noting that

$$\begin{aligned} P_{X_1,X_2,Y_1,Y_2}(x_1,x_2,y_1,y_2) &= P_{X_1,X_2}(x_1,x_2)P_{Y_1,Y_2,|X_1,X_2}(y_1,y_2|x_1,x_2) \\ &= P_{X_1,X_2}(x_1,x_2)P_{Y_1|X_1}(y_1|x_1)P_{Y_2|X_2}(y_2|x_2). \end{aligned}$$

I.e. we have a Markov chain! Thus

$$\begin{aligned} I(X_1,X_2;Y_1,Y_2) &= H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2) \\ &= H(Y_1,Y_2) - H(Y_1|X_1,X_2) - H(Y_2|X_1,X_2) \\ &= H(Y_1,Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &= I(X_1;Y_1) + I(X_2;Y_2). \end{aligned}$$

Maximizing for capacity yields

$$\begin{aligned} C = \max_{P_{X_1,X_2}} I(X_1, X_2; Y_1, Y_2) &\le \max_{P_{X_1,X_2}} \left\{ I(X_1;Y_1) + I(X_2;Y_2) \right\} \\ &= \max_{P_{X_1,X_2}} I(X_1;Y_1) + \max_{P_{X_1,X_2}} I(X_2;Y_2) \\ &= \max_{P_{X_1}} I(X_1;Y_1) + \max_{P_{X_2}} I(X_2;Y_2) \\ &= C_1 + C_2 \end{aligned}$$

The inequality can be made an equality by making $X_1 \perp\!\!\!\perp X_2$, In conclusion, setting $P^*_{X_1,X_2} = P^*_{X_1} \otimes P^*_{X_2}$ gives us $C_{max} = C_1 + C_2$.

■

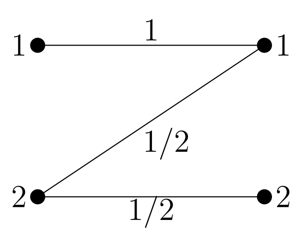
9. **The Z-channel.**
This channel has binary input and output alphabets and a channel transition kernel $P_{Y|X}$ as describe by the matrix $Q = \left[Q_{x,y}\right]_{x,y\in\{0,1\}}$:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix},$$

where $Q_{x,y} = P_{Y|X}(y|x)$, for all $x, y \in \{0, 1\}$.

(a) Draw a diagram of the Z-channel.

Solution.



■

(b) Find the capacity of the Z-channel and the maximizing input probability distribution.

Solution.
We obtain the mutual information of the input X with the output Y to be:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H_b(\mathbb{P}(Y=1)) - \mathbb{P}(X=0)\cdot 0 - \mathbb{P}(X=1)\cdot 1 \\ &= H_b\left(\frac{\mathbb{P}(X=1)}{2}\right) - \mathbb{P}(X=1). \end{aligned}$$

We note that this is a concave function with respect to $\mathbb{P}(X = 1)$. Thus

$$C = \max_{\mathbb{P}(X=1)} \left\{ H_b\left(\frac{\mathbb{P}(X = 1)}{2}\right) - \mathbb{P}(X = 1) \right\}.$$

Differentiating with respect to $\mathbb{P}(X = 1)$ we obtain $\mathbb{P}(X = 1) = 0.4$ to be the value at which our function attains its maximum. In conclusion,

$$C = H_b(0.4) - 0.4 = 0.322.$$

■