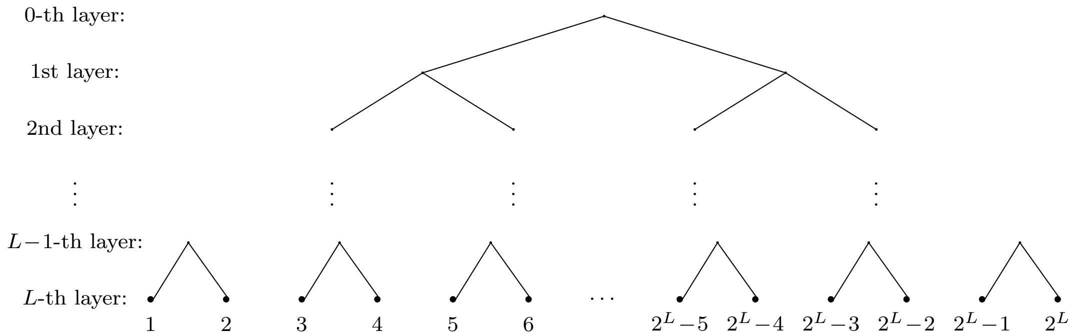
EE5139/EE6139
Solution 10

Semester I, AY2021/2022
Marco Tomamichel

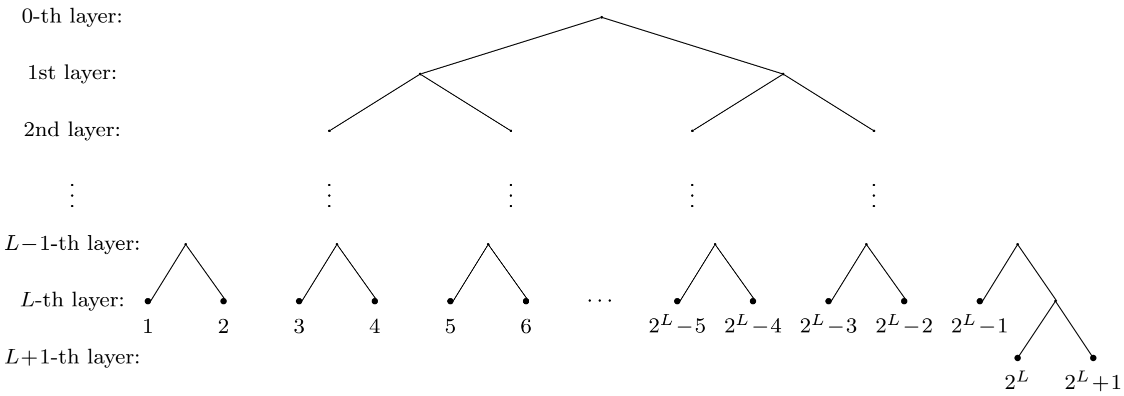
Choose any three (EE5139) or two (EE6139) of the below exercises. The remainder of the exercises will serve as a good preparation for the final exam.

Exercise 10.1 Binary Huffman Code

Suppose there is a source X with alphabet $\{1, 2, \ldots, n\}$ and each symbol has an equal probability. We use the Huffman algorithm to generate a binary code. Calculate the code length of each symbol and the average code length when $n = 2^L$ and $n = 2^L + 1$, where L is a positive integer.
Solution: For $n = 2^L$, the resultant tree of the Huffman algorithm looks like below:



The length of each codeword is L, and thus the average length is L.
For $n = 2^L + 1$, the resultant tree looks like this:



The length of the first 2^L-1 (actually arbitrarily picken 2^L-1) codewords is L, and the remaining 2 codewords are of length $L+1$. Thus, the average length is $\frac{2^L-1}{2^L+1} \cdot L + \frac{2}{2^L+1} \cdot (L+1) = L + \frac{2}{2^L+1}$.

Exercise 10.2 A Testing Problem

Let there be 9 visually identical balls. One of the balls is heavier than the rest, and the remaining 8 balls are of equal weight. The task is to find out the heavier ball with a balance. We let X be a random variable denoting the number of the heaviest ball, *i.e.*, $X = i$ if and only if the i-th ball is heaviest, Assume equal probability for each ball being the heavier one, *i.e.*, $p_X(i) = 1/9$ or each $i = 1, \ldots, 9$.

a.) Suppose, for the first weighing, you decide to weigh 3 balls against another 3 balls. Denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ (*i.e.*, the average uncertainty of X after observing Y)?

Solution: Observe that given any three outcome of Y, we will know the heaviest ball to be among some 3 balls. By the symmetric setup of this problem, the likelihood of the heaviest ball to be any of these 3 balls is same. Hence, $H(X|Y=y) = \log 3$ for each y. As a result, $H(X|Y) = \log 3$.

b.) Calculate the mutual information $I(X:Y) = H(X) - H(X|Y)$ and what is its interpretation in this context?

Solution: Since $H(X) = \log 9$, we have $I(X:Y) = H(X) - H(X|Y) = \log 3$. This means that the first weighing has led to a reduction of average uncertainty worth of $\log_2 3$-bits of information.

c.) Calculate $H(Y)$ and $H(Y|X)$ directly from their distributions. Using these term, recalculate the mutual information $I(X:Y) = H(Y) - H(Y|X)$. (Of course, you should end up with same result. Otherwise, check your answer.)

Solution: Note that $Y = y$ is equivalent to the event that the heaviest ball being among certain 3 balls. Thus, $p_Y(y) = 1/3$ for each y, and hence $H(Y) = \log 3$. On the other hand, given $X = x$, Y is deterministic. Thus, $H(Y|X = x) = 0$ for each x, and hence $H(Y|X) = 0$. In this case, $I(X:Y) = H(Y) - H(Y|X) = \log 3$.

d.) Suppose, instead, you decide to weigh 2 balls against another 2 balls for the first weighing, and denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ and hence, calculate $I(X:Y)$.

Solution: There are three cases:

- The balance tilted to the left: This occurs with probability $2/9$ when the heaviest ball is among certain 2 balls.
- The balance tilted to the right: This occurs with probability $2/9$ when the heaviest ball is among certain another 2 balls.
- The balance was leveled: This occurs with probability $5/9$ when the heaviest ball is among the remaining 5 balls.

Thus,

$$H(X|Y) = \frac{2}{9} \cdot \log 2 + \frac{2}{9} \cdot \log 2 + \frac{5}{9} \cdot \log 5 = \frac{4}{9} + \frac{5}{9} \cdot \log 5 = 1.7344 \text{ bits}.$$

Hence,

$$I(X:Y) = H(X) - H(X|Y) = 2 \cdot \log 3 - \frac{4}{9} - \frac{5}{9} \cdot \log 5 = 1.4355 \text{ bits}.$$

e.) Suppose, instead, you decide to weigh only one ball against another ball for the first weighing, and denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ and hence, calculate $I(X:Y)$.

Solution: Similar to the last step, we have

$$H(X|Y) = \frac{1}{9} \cdot \log 1 + \frac{1}{9} \cdot \log 1 + \frac{7}{9} \cdot \log 7 = \frac{7}{9} \cdot \log 7 = 2.1835 \text{ bits}.$$

Hence,

$$I(X:Y) = H(X) - H(X|Y) = 2 \cdot \log 3 - \frac{7}{9} \cdot \log 7 = 0.9864 \text{ bits}.$$

f.) Based on above steps, what is the optimal weighing strategy for the first weighing in order to maximize mutual information $I(X:Y)$?

Solution: The optimal strategy is to weigh balls in groups of 3.

g.) Suppose now you have narrowed down the candidate for the heavier ball to just 3 balls, following the argument above. What should be the optimal weighing pattern be for the second weighing?

Solution: By a similar argument, we should do a 1 ball against 1 ball weighing. Doing so will lead to $H(X|Y)$ of 0 since we will know with certainty which ball is heavier after the weighing. This will lead to the largest information gain since $H(X|Y) \geq 0$ and we have attained the minimum conditional entropy with this weighing.

Exercise 10.3 Another Testing Problem

Given 8 bottles of water, among which one of them is sugared. The likelihood of the x-th bottle to be sugared (prior to any testing) is described by the probability p_X. Suppose, for each $i \in \{1, \ldots, 8\}$, $p_X(i) = p_i$ where $(p_1, p_2, \ldots, p_8) = (\frac{7}{32}, \frac{6}{32}, \frac{6}{32}, \frac{5}{32}, \frac{3}{32}, \frac{3}{32}, \frac{1}{32}, \frac{1}{32})$. The task is to taste the water to figure out which bottle is sugared. (Obviously, you need to taste at most 7 times.)

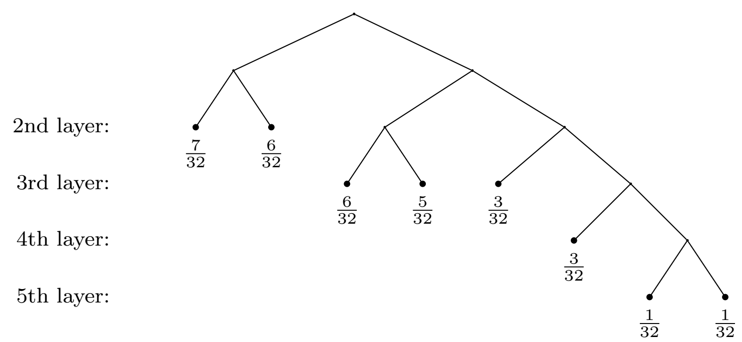
a.) Without mixing the water, what is the minimum expected number of tastes? In what order for tasting of the bottles, is this minimum expectation attained?

Solution: The best order is the descending order of the probability of sugaring, since the average number of taste is $\sum_{k=1}^{9} k \cdot p_{\pi_k}$. The minimum expectation is

$$\sum_{i=1}^{8} p_i l_i = 1 \cdot \frac{7}{32} + 2 \cdot \frac{6}{32} + 3 \cdot \frac{6}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{3}{32} + 6 \cdot \frac{3}{32} + 7 \cdot \frac{1}{32} + 7 \cdot \frac{1}{32} = 3.25.$$

b.) Allowing mixing, what is the strategy that minimizes the expected number of drinks that need to be tasted? What is the expected and maximal number of taste for this strategy? Which mixture should be tasted first? **Hint:** You may think about this problem as a coding problem.

Solution: Suppose we have a strategy that can identify the sugared bottle in finite steps. For each $x \in \{1, \ldots, 9\}$, one can use the strategy to generate a sequence of 0 and 1's by denoting down a 0 or 1 at i-th location if the i-th test is unsweet or sweet. This will result a uniquely decodable code since the strategy can identify x, and the average length of the code is the average number of test needed in this strategy. Knowing Huffman code to be optimal, we can construct a Huffman tree as follows (not unique)



The average length of above code (and thus the expected number of taste) is $\frac{90}{32}$. At first step, we may taste the mixture of the first two bottles.

Exercise 10.4 Joint Typicality

Let (X, Y) be a pair of random variables with X and Y distributed on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let (X^n, Y^n) be n i.i.d. copies of (X, Y). We define the set of joint typical sequences in $(\mathcal{X} \times \mathcal{Y})^n$

as (note that this is slightly different from Exercise 9.3)

$$\mathcal{A}_\epsilon^{(n)}(XY) := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \frac{1}{n} \log \frac{1}{p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y})} - H(XY) \right| \leq \epsilon \right\}.$$

A sequence $(\boldsymbol{x}, \boldsymbol{y})$ is said to be jointly typical if and only if it is in $\mathcal{A}_\epsilon^{(n)}(XY)$.

a.) Show that, for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}(XY)$, it holds that

$$2^{-n(H(X)+H(Y|X)+\epsilon)} \leq p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y}) \leq 2^{-n(H(X)+H(Y|X)-\epsilon)}.$$

Solution: By noticing that $H(X)+H(Y|X) = H(X,Y)$, it is straightforward to see above is just a rewriting of the definition of the joint typical set.

b.) For a typical $\boldsymbol{x} \in \mathcal{A}_\epsilon^{(n)}(X)$, a sequence $\boldsymbol{y} \in \mathcal{Y}^n$ is said to be *relatively typical* to $\boldsymbol{x}$ if and only if $(\boldsymbol{x}, \boldsymbol{y})$ is jointly typical. Given $\boldsymbol{y}$ to be relatively typical to $\boldsymbol{x}$, show that

$$2^{-n(H(Y|X)+2\epsilon)} \leq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}) \leq 2^{-n(H(Y|X)-2\epsilon)}.$$

Solution: Suppose $\boldsymbol{x}$ is ϵ-typical, and $\boldsymbol{y}$ is relative typical to $\boldsymbol{x}$. It must holds that,

$$\begin{array}{rcl} 2^{-n(H(X)+\epsilon)} \leq & p_{X^n}(\boldsymbol{x}) & \leq 2^{-n(H(X)-\epsilon)} \\ 2^{-n(H(X)+H(Y|X)+\epsilon)} \leq & p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y}) & \leq 2^{-n(H(X)+H(Y|X)-\epsilon)}. \end{array}$$

Thus, the value of $p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}) = p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y})/p_{X^n}(\boldsymbol{x})$ must be in the range between $2^{-n(H(X)+H(Y|X)+\epsilon)}/2^{-n(H(X)-\epsilon)}$ and $2^{-n(H(X)+H(Y|X)-\epsilon)}/2^{-n(H(X)+\epsilon)}$.

c.) Prove that the probability that Y^n is relatively typical to X^n tends to 1 as $n \to \infty$, namely,

$$p_{X^n,Y^n}(\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \boldsymbol{x} \in \mathcal{A}_\epsilon^{(n)}(X), (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}(XY)\}) \to 1$$

as $n \to \infty$.

Solution: Firstly, note that

$$\begin{aligned} 1 - \text{above probability} &= \Pr[X^n \notin \mathcal{A}_\epsilon^{(n)}(X) \vee Y^n \notin \mathcal{A}_\epsilon^{(n)}(XY)] \\ &\leq \Pr[X^n \notin \mathcal{A}_\epsilon^{(n)}(X)] + \Pr[X^nY^n \notin \mathcal{A}_\epsilon^{(n)}(XY)]. \end{aligned}$$

However, we have already shown that $\Pr[X^n \notin \mathcal{A}_\epsilon^{(n)}(X)] \to 0$ as $n \to \infty$, and it follows from exactly same proof that $\Pr[X^nY^n \notin \mathcal{A}_\epsilon^{(n)}(XY)] \to 0$ as $n \to \infty$, which finished the proof.

Exercise 10.5 Information Spectrum Analysis

Let $p_{Y|X}$ be a channel from input set $\mathcal{X}$ to output set $\mathcal{Y}$. $\mathcal{X}$ and $\mathcal{Y}$ need not be discrete.

a.) Suppose we use the channel *once*. Show that there exist a code with M codewords with average probability of error ϵ satisfying

$$\epsilon \leq \Pr\left[\log \frac{p_{Y|X}(Y|X)}{p_Y(Y)} \leq \log M + \gamma\right] + 2^{-\gamma}.$$

for any choice of $\gamma > 0$ and any input distribution P_X where $P_Y(y) = \sum_x P_{Y|X}(y|x)P_X(x)$.
Hint: Generate codewords independently according to P_X. Instead of using typical set for decoding, use $\hat{m} \in \{1, \ldots, M\}$ as the transmitted message if it is the unique one satisfying

$$\log \frac{P_{Y|X}(y|x(\hat{m}))}{P_Y(y)} \geq \log M + \gamma.$$

If there is no unique $\hat{m}$ satisfying the above condition, declare an error. The analysis to arrive at the one-shot (finite blocklength) bound above is very similar to typical set decoding. A stronger version of this bound (for maximum error) was shown by Feinstein.

Solution: As provided in the hint, we generate M codewords $x(m)$ independently from P_X. To send message m, we transmit codeword $x(m)$. Decode using the rule given above. Without loss of generality, we assume $m = 1$ was sent. An error occurs if and only if one or more of the following events occurs:

$$\mathcal{E}_1 := \left\{ \log \frac{P_{Y|X}(Y|X(1))}{P_Y(Y)} < \log M + \gamma \right\}$$
$$\mathcal{E}_2 := \left\{ \exists\, \tilde{m} \neq 1 : \log \frac{P_{Y|X}(Y|X(\tilde{m}))}{P_Y(Y)} \geq \log M + \gamma \right\}$$

The probability of error can be bounded as

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2)$$

Now note that $(X(1), Y) \sim P_X P_{Y|X}$, and thus $\Pr(\mathcal{E}_1)$ gives the first term in the bound we have to show. We simply have to show that $\Pr(\mathcal{E}_2) \leq 2^{-n\gamma}$. For this, consider

$$\begin{aligned}
\Pr(\mathcal{E}_2) &= \Pr\left(\exists\, \tilde{m} \neq 1 : \log \frac{P_{Y|X}(Y|X(\tilde{m}))}{P_Y(Y)} \geq \log M + \gamma \right) \\
&\overset{(a)}{\leq} \sum_{\tilde{m}=2}^{M} \Pr\left(\log \frac{P_{Y|X}(Y|X(\tilde{m}))}{P_Y(Y)} \geq \log M + \gamma \right) \\
&\overset{(b)}{=} \sum_{\tilde{m}=2}^{M} \sum_{x,y} P_X(x) P_Y(y) \mathbf{1}\left\{ \log \frac{P_{Y|X}(y|x)}{P_Y(y)} \geq \log M + \gamma \right\} \\
&\overset{(c)}{\leq} \sum_{\tilde{m}=2}^{M} \sum_{x,y} P_X(x) P_{Y|X}(y|x) M^{-1} 2^{-\gamma} \mathbf{1}\left\{ \log \frac{P_{Y|X}(y|x)}{P_Y(y)} \geq \log M + \gamma \right\} \\
&\overset{(d)}{\leq} \sum_{\tilde{m}=2}^{M} \sum_{x,y} P_X(x) P_{Y|X}(y|x) M^{-1} 2^{-\gamma} \\
&\overset{(e)}{\leq} 2^{-\gamma}
\end{aligned}$$

where (a) is due to the union bound, and (b) is due to the fact that for $\tilde{m} \neq 1$, the codeword $X(\tilde{m})$ and channel output Y are independent, and (c) is due to the fact that we're only summing over all (x, y) such that $\log \frac{P_{Y|X}(y|x)}{P_Y(y)} \geq \log M + \gamma$, and for (d) we dropped the indicator, and for (e) we use the fact that $\sum_{x,y} P_X(x) P_{Y|X}(y|x) = 1$ and that there are $M - 1$ terms in the outer sum.

b.) Based on part (a), prove the channel coding theorem for finite $\mathcal{X}, \mathcal{Y}$ and memoryless channels. **Hint:** Set P_{X^n} above to be the n-fold product distribution corresponding to a capacity-achieving input distribution $P_X \in \arg\max_{P_X} I(X;Y)$. Set γ above to be $n\gamma'$ for some $\gamma' > 0$. Set $\log M = n(C - 2\gamma')$. Apply the law of large numbers to the first term to see that there exists a sequence of $(n, 2^{n(C-2\gamma')})$-codes with vanishing average error probabilities.

Solution: Going to the n-fold (n channel uses setting), we have that there exists a code with blocklength n, M_n codewords and average error probability ϵ_n satisfying

$$\epsilon_n \leq \Pr\left(\log \frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \log M_n + \gamma \right) + 2^{-\gamma}.$$

Choose P_{X^n} to be the n-fold product distribution corresponding to a capacity-achieving input distribution $P_X \in \arg\max_{P_X} I(X;Y)$. Since the channel is a DMC, we have

$$\begin{aligned}
\Pr\left(\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \log M_n + \gamma\right) &= \Pr\left(\sum_{i=1}^n \log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} \le \log M_n + n\gamma'\right)\\
&= \Pr\left(\sum_{i=1}^n \log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} \le n(C-2\gamma') + n\gamma'\right)\\
&= \Pr\left(\frac{1}{n}\sum_{i=1}^n \log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} \le C-\gamma'\right)
\end{aligned}$$

Since

$$\mathbb{E}\left[\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)}\right] = I(X;Y) = C$$

for all i, we have that the probability above tends to zero by the weak law of large numbers. Clearly, the second term in the bound $2^{-\gamma} = 2^{-n\gamma'}$ also tends to zero because $\gamma' > 0$. So we have demonstrated a sequence of codes for which $\epsilon_n \to 0$ and the code rate is $C - 2\gamma'$ which is arbitrarily close to C.

c.) Again consider the setup in (b). Let

$$V := \mathrm{Var}\left(\log\frac{P^\star_{Y|X}(Y|X)}{P^\star_Y(Y)}\right)$$

be evaluated at a capacity-achieving input distribution. Based on part (a), show that, by the central limit theorem, there exists a sequence of codes indexed by blocklength n, with sizes M_n satisfying

$$\log M_n = nC + \sqrt{nV}\Phi^{-1}(\epsilon) + o(\sqrt{n})$$

such that the average error probability is no larger than $\epsilon + o(1)$.

Solution: Now we again use the bound

$$\epsilon_n \le \Pr\left(\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \log M_n + n\gamma'\right) + 2^{-n\gamma'}.$$

with $\gamma' = \frac{\log n}{n}$ so the final term is $1/n$. Plug the value of M_n into the probability in the bound above. We have

$$\begin{aligned}
&\Pr\left(\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \log M_n + n\gamma'\right)\\
&= \Pr\left(\sum_{i=1}^n \log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} \le nC + \sqrt{nV}\Phi^{-1}(\epsilon) + \log n\right)\\
&= \Pr\left(\frac{1}{\sqrt{nV}}\sum_{i=1}^n \left(\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} - C\right) \le \Phi^{-1}(\epsilon) + O\left(\frac{\log n}{\sqrt{nV}}\right)\right)
\end{aligned}$$

Now note that for all $i \in [n]$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{V}}\left(\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} - C\right)\right] &= 0\\
\mathrm{Var}\left[\frac{1}{\sqrt{V}}\left(\log\frac{P_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} - C\right)\right] &= 1
\end{aligned}$$

so the random variable in the probability converges to a standard Gaussian by the central limit theorem. Consequently,

$$\Pr\left(\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \log M_n + n\gamma'\right) \to \epsilon$$

and we are done.

Exercise 10.6 Channel Coding and Capacity

Consider that there is a binary symmetric channel (BSC) with crossover probability ϵ. We use a coding scheme on this channel that encodes messages a_1 and a_2 as 000 and 111, respectively. On the other hand, we use a decoding scheme that relies on the majority principle, *i.e.*, $001 \mapsto 0$ and $011 \mapsto 1$.

a.) Merge the above encoder, the BSC, and the above decoder into a new channel. Describe the input/output alphabets and the channel rule (*i.e.*, the conditional probability of outputs given inputs) of this new channel.

Solution: Both The input and output alphabets are still $\{0, 1\}$. Denoting the original BSC by W, and the new channel by $\tilde{W}$, we have

$$\begin{aligned}
\tilde{W}(0|0) &= W^{\otimes 3}(000|000) + W^{\otimes 3}(001|000) + W^{\otimes 3}(010|000) + W^{\otimes 3}(100|000) \\
&= (1-\epsilon)^3 + 3\epsilon(1-\epsilon)^2 = 1 - 3\epsilon^2 + 2\epsilon^3, \\
\tilde{W}(1|0) &= 1 - \tilde{W}(0|0) = 3\epsilon^2 - 2\epsilon^3,
\end{aligned}$$

and by symmetry $\tilde{W}(1|1) = 1 - 3\epsilon^2 + 2\epsilon^3$ and $\tilde{W}(0|1) = 3\epsilon^2 - 2\epsilon^3$. In other words, the resultant channel is a BSC with crossover probability $3\epsilon^2 - 2\epsilon^3$.

b.) What is the capacity of the new channel if the crossover probability of the original BSC $\epsilon = 0.1$?

Solution: Let us denote $h(\epsilon) := \epsilon \log \frac{1}{\epsilon} + (1-\epsilon)\log\frac{1}{1-\epsilon}$. The crossover probability of the new channel is $\hat{\epsilon} = 0.028$. The channel capacity per channel use is given by

$$\begin{aligned}
C &= \max I(\hat{M} : M) \\
&= \max H(\hat{M}) - \sum_{m \in \{a_1, a_2\}} P_M(m) H(\hat{M}|M = m) \\
&= \max H(\hat{M}) - h(\hat{\epsilon}) \\
&= 1 - h(0.028) \\
&= 0.81574 \text{ bits per channel use.}
\end{aligned}$$

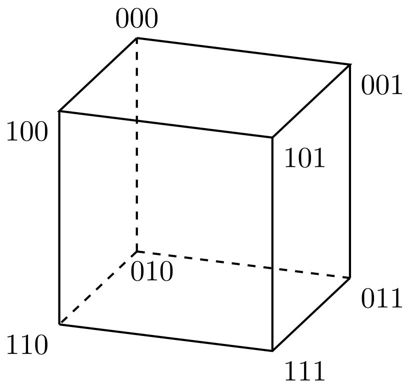
The code rate is $\frac{1}{3}$. Thus, the channel capacity in bits per transmission is $\frac{C}{3} = 0.27191$ bits per transmission.

c.) What is the capacity of the original BSC channel with $\epsilon = 0.1$?

Solution: The capacity of the original channel is simply $1 - h(0.1) = 0.531004$.

d.) Suppose that we have 4 input messages (each to be sent with equal probability) instead. Find the best code of length 3 that minimizes the average decoding error.

Solution: To find the best codewords of four messages with length 3, we should let the codewords be as far as possible.



From the codewords cube, we can see that one possible optimal set is $\{000, 101, 011, 110\}$. Each of these codewords has two different bits comparing to other codewords. The decoding rule is to assign received sequences to the nearest codewords. Hence,

$$P(rec = trans) = 0.9 \times 0.9 \times 0.9 + 0.9 \times 0.9 \times 0.1 = 0.81.$$

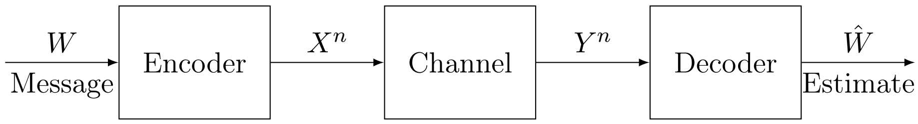
The probability of error is 0.19.

e.) Consider a binary erasure channel (BEC) with erasure probability $\epsilon = 0.1$. Suppose we encode the messages a_1 and a_2 as 000 and 111, respectively. On the decoder side, if we receive $\perp\perp\perp$, we will randomly assign it to 000 or 111 with equal probability; otherwise, we decode according to the survived symbol, *i.e.*, $0 \perp\perp \mapsto a_1$ and $\perp\perp 1 \mapsto a_2$. What is the average decoding error in this case?

Solution: The received sequences $00\perp$, $0\perp0$, $\perp00$, $\perp\perp0$, $\perp0\perp$, $0\perp\perp$, 000 can be decoded as 0. The received sequences $11\perp$, $1\perp1$, $\perp11$, $\perp\perp1$, $\perp1\perp$, $1\perp\perp$, 111 can be decoded as 1. Only the $\perp\perp\perp$ may cause the error. The probability of receiving $\perp\perp\perp$ is 0.1^3. Thus, the probability of error for this code is $0.1^3/2 = 0.0005$.

f.) Merge the encoder (from step e.)), the BEC, and the decoder (from step e.)) into another new channel. Show, without computation, the capacity of the new channel is no larger than three times the capacity of the BEC.

Solution:



By the data processing inequality,

$$I(W : \hat{W}) \le I(X^n : Y^n),$$

and thus

$$C_W = \frac{1}{n} \max_{p(w)} I(W : \hat{W}) \le \frac{1}{n} \max_{p(x^n)} I(X^n : Y^n) \le C.$$

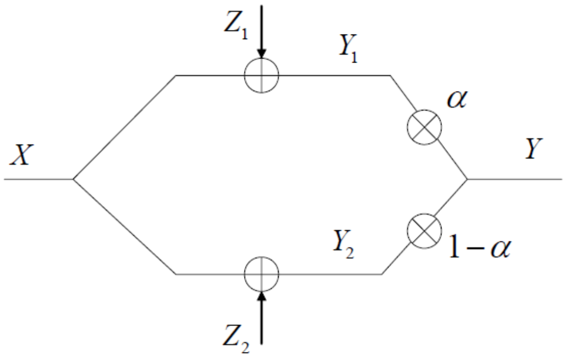
Hence, the capacity of channel per transmission is not increased by merging the encoder and decoder.

Exercise 10.7 A Gaussian channel

Consider a Gaussian channel shown below, in which the transmitted signal X with $\mathbb{E}[X^2] = P$ is received by two antennas with $Y_1 = X + Z_1$ and $Y_2 = X + Z_2$ where Z_1 and Z_2 are independent with $\mathbb{E}[Z_i^2] = \sigma_i^2$ ($\sigma_1^2 < \sigma_2^2$). The signals at the two antennas are combined as

$$Y = \alpha Y_1 + (1 - \alpha) Y_2$$

before decoding, where $0 \le \alpha \le 1$.



a.) Find the capacity of this channel for a given α. Please provide the units of capacity.

Solution: This is a Gaussian channel with signal power P and noise power

$$\alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2.$$

Hence, the capacity is

$$\frac{1}{2}\log\left(1 + \frac{P}{\alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2}\right) \quad \text{bits per channel use.}$$

b.) Using your result at the previous point, find the optimal α that maximizes the capacity and write down the corresponding maximum capacity in terms of P, σ_1^2 and σ_2^2.

Solution: $\alpha^s tar = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. Maximum capacity $C^s tar = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2} + \frac{P}{\sigma_2^2}\right)$ bits per channel use.