Handout 19 | Principles of Digital Communications
Solution to Midterm exam | Apr. 27, 2022

4 problems, 165 minutes.
1 sheet (2 pages) of notes allowed.

Good Luck!

PLEASE WRITE YOUR NAME ON EACH SHEET OF YOUR ANSWERS.

PLEASE WRITE THE SOLUTION OF EACH PROBLEM ON A SEPARATE SHEET.

Problem 1. Consider the following conditional probability distribution,

$$P_{Y|H}(y\,|-1) = \begin{cases} 0.2 & y=2 \\ 0.2 & y=1 \\ 0.3 & y=-1 \\ 0.3 & y=-2 \end{cases} \qquad P_{Y|H}(y\,|\,1) = \begin{cases} 0.3 & y=2 \\ 0.3 & y=1 \\ 0.21 & y=-1 \\ 0.19 & y=-2 \end{cases}$$

Assume that $P_H(-1) = P_H(1) = 1/2$. We consider a hypothesis problem of estimating H after observing Y. Determine whether the following statistics are sufficient statistics or not. Provide a justification for your answer.

a) (2 points) $T_1(Y) = \text{argmax}_h P_{Y|H}(Y\,|\,h)$.

It is not a sufficient statistic. For example, $T_1(-1) = -1 = T_1(-2) = -1$, *however,*

$$\frac{P_{H|Y}(-1|-1)}{P_{H|Y}(1|-1)} \neq \frac{P_{H|Y}(-1|-2)}{P_{H|Y}(1|-2)}.$$

If $T_1(.)$ *is a sufficient statistic, then likelihood of* y*'s with similar sufficient statistic should have been equal.*

b) (2 points) $T_2(Y) = \Big(T_1(Y), P_{Y|H}\big(Y\,|\,T_1(Y)\big)\Big)$, i.e. a tuple composed of $T_1(Y)$ and the likelihood of observing Y conditioned on $H = T_1(Y)$.

It is still not a sufficient statistic. Note that $T_2(-1) = T_2(-2) = (-1, 0.3)$. *However, as we have saw in (a), the likelihood ratio of these two* y*'s are not equal.*

c) (4 points) $T_3(Y) = P_{Y|H}\big(Y\,|\,T_1(Y)\big)/P_{Y|H}\big(Y\,|-T_1(Y)\big)$.

Interestingly enough, this is a sufficient statistic. We can see that there are three possible value of $T_3(y)$, *namely* $\{30/19, 3/2, 30/21\}$. *Each of these possible values corresponds to unique likelihood ratio.*

Now, we introduce another hypothesis with conditional probability distribution,

$$P_{Y|H}(y\,|\,0) = \begin{cases} 0.27 & y=2 \\ 0.26 & y=1 \\ 0.24 & y=-1 \\ 0.23 & y=-2 \end{cases}$$

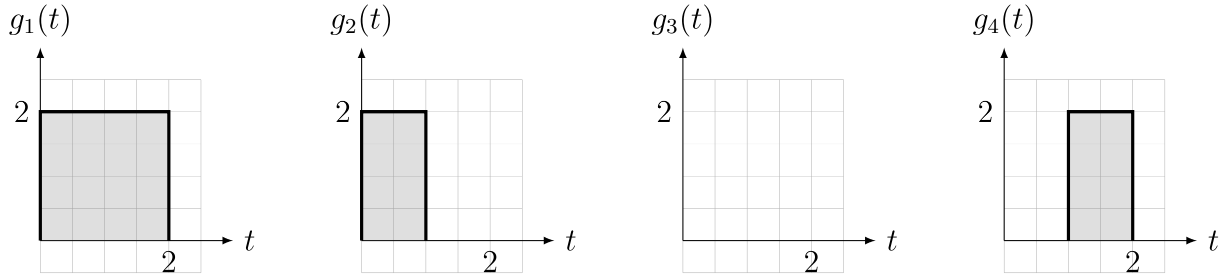
with $P_H(0) = P_H(-1) = P_H(1) = 1/3$.

d) (4 points) Is $T_3(Y)$ a sufficient statistic? Provide a justification for your answer.

The addition of this hypothesis renders the statistics to be insufficient. We have $T_3(1) = T_3(2)$. *However,*

$$\big(P_{H|Y}(1|1), P_{H|Y}(0|1), P_{H|Y}(-1|1)\big) \neq \big(P_{H|Y}(1|2), P_{H|Y}(0|2), P_{H|Y}(-1|2)\big).$$

PROBLEM 2. Consider the following waveforms,



a) (2 points) Find an orthonormal basis $\psi_1(t)$ and $\psi_2(t)$ for these waveforms such that the support of these two basis waveforms are disjoint.

An orthonormal basis function is given by,

$$\psi_1(t) = \mathbb{1}\{t \in [0,1]\} \qquad \psi_2(t) = \mathbb{1}\{t \in (1,2]\}.$$

One can parametrize the waveforms as

$$g_1(t) = 2\psi_1(t) + 2\psi_2(t) \qquad g_2(t) = 2\psi_1(t)$$
$$g_3(t) = 0 \qquad g_4(t) = 2\psi_2(t).$$

b) (2 points) Do these waveforms form a minimal energy signal set? If it is not, give a minimal energy set by applying a translation.

One can take the averages of these waveforms

$$\bar{g}(t) = \frac{1}{4}\sum_{i=1}^{4} g_i(t) = \psi_1(t) + \psi_2(t).$$

Observe that the average is not a zero function. Hence these waveforms are not energy minimal. We can subtract the average waveform from the original waveforms to obtain the minimal energy set,

$$\tilde{g}_1(t) = \psi_1(t) + \psi_2(t) \qquad \tilde{g}_2(t) = \psi_1(t) - \psi_2(t)$$
$$\tilde{g}_3(t) = -\psi_1(t) - \psi_2(t) \qquad \tilde{g}_4(t) = -\psi_1(t) + \psi_2(t).$$

We will denote the waveforms corresponding to the minimal energy signal set as $\tilde{g}_i(t)$. Let us assume that the received waveform $R(t)$ is given by $R(t) = \tilde{g}_i(t) + N(t)$, where i is the transmitted message and $N(t)$ is white Gaussian noise of intensity $N_0/2 = 1$. Let $Y_1 = \langle R(t), \psi_1(t)\rangle$ and $Y_2 = \langle R(t), \psi_2(t)\rangle$. Let us assume that all messages are equiprobable.

c) (4 points) Conditioned on i-th message is being sent, what is the distribution of (Y_1, Y_2).

We have,

$$Y_1 = \langle R(t), \psi_1(t)\rangle = \langle \tilde{g}_i(t), \psi_1(t)\rangle + \langle N(t), \psi_1(t)\rangle \sim N(\langle \tilde{g}_i(t), \psi_1(t)\rangle, 1)$$
$$Y_2 = \langle R(t), \psi_2(t)\rangle = \langle \tilde{g}_i(t), \psi_2(t)\rangle + \langle N(t), \psi_2(t)\rangle \sim N(\langle \tilde{g}_i(t), \psi_2(t)\rangle, 1).$$

d) (4 points) What is the MAP decision region and the corresponding error probability of this transmission scheme?

Let us write

$$Y_1 = X_1 + Z_1$$
$$Y_2 = X_2 + Z_2$$

where $X_1 = \langle \tilde{g}_H(t)\psi_1(t)\rangle$ *and* $X_2 = \langle \tilde{g}_H(t)\psi_2(t)\rangle$ *with* H *is the true message. We have* $Var(Z_1) = N_0 \langle \psi_1(t), \psi_1(t)\rangle /2 = 1$.

We note that this is the case of $4QAM$*, the optimal decision region corresponds to each quadrant, i.e.,* $D_1 = \{(y_1, y_2) : y_1 > 0, y_2 > 0\}$,$D_2 = \{(y_1, y_2) : y_1 > 0, y_2 < 0\}$,$D_3 = \{(y_1, y_2) : y_1 < 0, y_2 < 0\}$ *and* $D_4 = \{(y_1, y_2) : y_1 < 0, y_2 > 0\}$.

Due to symmetry, it is enough to study the error probability for one waveforms. We take the case where $H = 1$*. Here, the error event is given by*

$$\Pr(Error) = \Pr(Z_1 < -1 \cup Z_2 < -1) = 2Q(1) - Q(1)^2.$$

Now, we will assume that the noise process $N(t)$ is Gaussian but not white, and, for all t,

$$\mathbb{E}\big[N(t)N(t-\tau)\big] = \begin{cases} 1 - |\tau| & |\tau| < 1 \\ 0 & \text{otherwise.} \end{cases}.$$

The receiver does not know of this new noise model, hence it still computes (Y_1, Y_2) and forms its decision based on the decision region that you have developed in d).

e) (4 points) Is (Y_1, Y_2) a sufficient statistic?

It is not a sufficient statistic. Let us consider another statistics $Y_0 = \langle R(t), \mathbb{1}\{t \in [-1, 0)\}\rangle$*. We can see that the correlation* $E[Y_0(Y_1 - E[Y_1])] \neq 0$*, while* $E[Y_0(Y_2 - E[Y_2])] = 0$*. Hence,* Y_0 *contains information about the noise component of* Y_1 *which is not contained in* Y_2*. This imply that we can reduce the noise of* Y_1 *by using* Y_0*. Let us define*

$$\tilde{Y_1} = Y_1 - \frac{1}{2}Y_0$$

The covariance matrix of $(\tilde{Y_1}, Y_2)$ *is given by*

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{bmatrix}$$

This covariance matrix is better than the covariance matrix given in the next point. The fact that we can reduce the noise by adding another statistic imply that Y_1, Y_2 *are not sufficient statistics.*

f) (4 points) What is the error probability of the receiver under this new noise model? It is sufficient to give express the error probability as an integral with an explicit integration region.

The receiver uses the same decision region as in (d). But now, Z_1 *and* Z_2 *are not independent. In this case*

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma\right)$$

with

$$\Sigma = \begin{bmatrix} \alpha & \rho \\ \rho & \alpha \end{bmatrix}$$

where

$$\begin{aligned}
\alpha &= E[Z_1 Z_1] \\
&= E\left[\iint N(\tau_1)N(\tau_2)\psi_1(\tau_1)\psi_1(\tau_2)d\tau_1 d\tau_2\right] \\
&= \int_0^1 \int_{-\tau_1}^{1-\tau_1} E[N(\tau_1)N(\tau_1+s)]d\tau_1 ds \\
&= \int_0^1 \frac{\tau_1(2-\tau_1)}{2} + \frac{(1-\tau_1)(1+\tau_1)}{2} d\tau_1 \\
&= \frac{1}{2} - \frac{1}{6} + \frac{1}{2} - \frac{1}{6} \\
&= \frac{2}{3}
\end{aligned}$$

and ρ,

$$\begin{aligned}
\rho &= E\left[\iint N(\tau_1)N(\tau_2)\psi_1(\tau_1)\psi_2(\tau_2)d\tau_1 d\tau_2\right] \\
&= \int_0^1 \int_{1-\tau_1}^{1} E[N(\tau_1)N(\tau_1+s)]d\tau_1 ds \\
&= \int_0^1 \frac{\tau_1^2}{2} d\tau_1 \\
&= \frac{1}{6}.
\end{aligned}$$

Due to the noise not being white, we have to consider two classes of waveform, $\{\tilde{g}_1, \tilde{g}_3\}$ and $\{\tilde{g}_2, \tilde{g}_4\}$. Hence the probability of error is given by,

$$\begin{aligned}
P(Error) = 1 &- \frac{1}{2}\int_{-1}^{\infty}\int_{-1}^{\infty} \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\left[y_1, y_2\right]\Sigma^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) dy_1 dy_2 \\
&- \frac{1}{2}\int_{-\infty}^{1}\int_{-1}^{\infty} \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\left[y_1, y_2\right]\Sigma^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) dy_1 dy_2
\end{aligned}$$

PROBLEM 3. Consider a binary hypothesis testing problem with the hypothesis $H \in \{0,1\}$ and the observation $Y \in \mathcal{Y}$. Unlike the standard decision problem where the decoder has to produce and estimate $\hat{H} \in \{0,1\}$, in this problem decoder can produce an estimate $\hat{H} \in \{0,1,?\}$. The symbol ? gives the decoder the option to say "I don't know". Let D_0, D_1, and $D_?$ denote the subsets of observation space $\mathcal{Y}$ for which the decoder declares $0, 1$ and ? respectively.

a) (4 points) We say the decoder makes an error if $\hat{H} \neq ?$ and $\hat{H} \neq H$. Show that the probability of error, P_e, equals

$$P_e = \sum_{y \in D_1} p_H(0) p_{Y|H}(y|0) + \sum_{y \in D_0} p_H(1) p_{Y|H}(y|1).$$

We have

$$\begin{aligned} P_e &= \Pr(Y \in D_0, H = 1) + \Pr(Y \in D_1, H = 0) \\ &= \Pr(Y \in D_0 | H = 1) p_H(1) + \Pr(Y \in D_1 | H = 0) p_H(0) \\ &= \sum_{y \in D_0} p_{Y|H}(y|1) p_H(1) + \sum_{y \in D_1} p_{Y|H}(y|0) p_H(0). \end{aligned}$$

b) (4 points) Show that the probability that $\hat{H} = ?$ equals

$$P_? = \sum_{y \in D_?} p_H(0) p_{Y|H}(y|0) + p_H(1) p_{Y|H}(y|1).$$

We have,

$$\begin{aligned} P_? &= \Pr(Y \in D_?) \\ &= \Pr(Y \in D_?, H = 1) + \Pr(Y \in D_?, H = 0) \\ &= \Pr(Y \in D_? | H = 1) p_H(1) + \Pr(Y \in D_? | H = 0) p_H(0) \\ &= \sum_{y \in D_?} p_{Y|H}(y|1) p_H(1) + p_{Y|H}(y|0) p_H(0). \end{aligned}$$

c) (4 points) Suppose there is a unit cost for making an error (in the sense of (a)) and a cost of $c_? \geq 0$ for declaring "I don't know". Find the decision rule that minimizes the expected cost (i.e., $P_e + c_? P_?$) and express it in terms of the a posteriori probabilities of the hypotheses.

To make the dependence on Y explicit, let us denote the decision rule as $\hat{H}(y)$. We have

$$\begin{aligned} &P_e + c_? P_? \\ &= \sum_y p_Y(y) \Big[\mathbb{1}\{\hat{H}(y) = 0\} p_{H|Y}(1|y) + \mathbb{1}\{\hat{H}(y) = 1\} p_{H|Y}(0|y) \\ &\qquad\qquad + \mathbb{1}\{\hat{H}(y) = ?\} c_? \big[P_{H|Y}(1|y) + P_{H|Y}(0|y) \big] \Big]. \end{aligned}$$

Therefore, to minimize the expected cost, we need to choose $H(y)$ to minimize the terms in the inner bracket. This leads to the following conditions

- $H(y) = 0$ *if* $p_{H|Y}(1|Y) < p_{H|Y}(0|Y)$ *and* $p_{H|Y}(1|Y) < c_?[p_{H|Y}(0|Y) + p_{H|Y}(1|Y)]$. *Given that* $c_? > 0$*, this condition is equivalent to* $\frac{p_{H|Y}(1|Y)}{p_{H|Y}(0|Y)} < \min\{1, c_?/(1-c_?)\}$.
- $H(y) = 1$ *if* $p_{H|Y}(0|Y) < p_{H|Y}(1|Y)$ *and* $p_{H|Y}(0|Y) < c_?[p_{H|Y}(0|Y) + p_{H|Y}(1|Y)]$. *Given that* $c_? > 0$*, this condition is equivalent to* $\frac{p_{H|Y}(1|Y)}{p_{H|Y}(0|Y)} > \max\{1, (1-c_?)/c_?\}$.
- $H(y) =?$ otherwise.

To summarize,

$$H(y) = \begin{cases} 0 & \frac{p_{H|Y}(1|Y)}{p_{H|Y}(0|Y)} \leq \min\{1, c_?/(1-c_?)\} \\ 1 & \max\{1, (1-c_?)/c_?\} \geq \frac{p_{H|Y}(1|Y)}{p_{H|Y}(0|Y)} \\ ? & \text{otherwise.} \end{cases}$$

Note that this is equivalent to

$$H(y) = \begin{cases} ? & \frac{p_{H|Y}(1|Y)}{p_{H|Y}(0|Y)} \in \left[\frac{c_?}{1-c_?}, \frac{1-c_?}{c_?}\right] \\ H_{MAP}(y) & \text{otherwise.} \end{cases}$$

d) (4 points) Consider now the the m-ary case, with costs as in (c). Show that the rule that that minimizes the expected cost is of the form

$$\hat{H}(y) = \begin{cases} ? & \text{(condition)} \\ \hat{H}_{MAP}(y) & \text{otherwise} \end{cases}$$

and determine the appropriate expression for (condition).

We only need to reconsider the case where $H(y) =?$*. We need the following condition to hold for all possible hypothesis* i

$$c_? \sum_{j=1}^{m} p_{H|Y}(j|y) < \sum_{\substack{j=1 \\ j \neq i}}^{m} p_{H|Y}(i|y).$$

By the fact that $\sum_{i=1}^{m} p_{H|Y}(i|y) = 1$*, this can also be written as,*

$$c_? p_{H|Y}(i|y) \leq (1-c_?)(1-p_{H|Y}(i|y)).$$

So we have that the required condition is equal to,

$$\frac{p_{H|Y}(i|y)}{1-p_{H|Y}(i|y)} \leq \frac{1-c_?}{c_?}$$

for all i.

PROBLEM 4. Consider a binary hypothesis testing problem where the observation Y has conditional distribution $p_{Y|H}$. Suppose that we use the MAP rule to decide on $\hat{H}$.

a) (4 points) Show that

$$\Pr(error) = \sum_y \min\{p_H(0)p(y|0),\, p_H(1)p(y|1)\}.$$

We have,

$$\Pr(Error) = \sum_y p_Y(y)\Big[\mathbb{1}\{\hat{H}(y) = 0\}p_{H|Y}(1|y) + \mathbb{1}\{\hat{H}(y) = 1\}p_{H|Y}(0|y)\Big].$$

We recall that the MAP rule is such that

$$\hat{H}(y) = \begin{cases} 0 & p_{H|Y}(1|y) \le p_{H|Y}(0|y) \\ 1 & p_{H|Y}(1|y) > p_{H|Y}(0|y) \end{cases}$$

such that

$$\begin{aligned} \Pr(Error) &= \sum_y \min\{p_{HY}(1, y), p_{HY}(0, y)\} \\ &= \sum_y \min\{p_{Y|H}(y|1)P_H(1), p_{Y|H}(y|0)P_H(0)\}. \end{aligned}$$

For $s \in [0, 1]$ define $\mu(s) := \sum_y p_{Y|H}(y|0)^{1-s} p_{Y|H}(y|1)^s$.

b) (4 points) Show that for any $s \in [0, 1]$

$$\Pr(error) \le P_H(0)^{1-s}P_H(1)^s\mu(s).$$

Hint: for non-negative a, b, and any $s \in [0, 1], \min\{a, b\} \le a^{1-s}b^s$.

Utilizing the hint, we have

$$\begin{aligned} \Pr(Error) &= \sum_y \min\{p_{Y|H}(y|0)P_H(0), p_{Y|H}(y|1)P_H(1)\} \\ &\le \sum_y \big(p_{Y|H}(y|0)P_H(0)\big)^{1-s}\big(p_{Y|H}(y|1)P_H(1)\big)^s \\ &= P_H(0)^{1-s}P_H(1)^s\mu(s). \end{aligned}$$

c) (4 points) Show that $\frac{d\mu(s)}{ds} = \sum_y p_{Y|H}(y|0)^{1-s}p_{Y|H}(y|1)^s\Lambda(y)$

where $\Lambda(y) = \ln[p_{Y|H}(y|1)/p_{Y|H}(y|0)]$, and $\frac{d^2\mu(s)}{ds^2} \ge 0$.

Notice that we can write,

$$\begin{aligned} \mu(s) &= \sum_y p_{Y|H}(y|0)\left(\frac{p_{Y|H}(y|1)}{p_{Y|H}(y|0)}\right)^s \\ &= \sum_y p_{Y|H}(y|0)\exp(s\Lambda(y)). \end{aligned}$$

Hence, the first derivative is given by

$$\begin{aligned}\frac{d\mu(s)}{ds} &= \sum_y p_{Y|H}(y|0)\frac{d\exp(s\Lambda(y))}{ds}\\ &= \sum_y p_{Y|H}(y|0)\exp(s\Lambda(y))\Lambda(y)\\ &= \sum_y p_{Y|H}(y|0)^{1-s}p_{Y|H}(y|1)^s\Lambda(y)\end{aligned}$$

We also have

$$\begin{aligned}\frac{d^2\mu(s)}{ds^2} &= \sum_y p_{Y|H}(y|0)\Lambda(y)\frac{d\exp(s\Lambda(y))}{ds}\\ &= \sum_y p_{Y|H}(y|0)\Lambda(y)^2\exp(s\Lambda(y)).\end{aligned}$$

This function is always non-negative as $\Lambda(y)^2$ is non-negative, $p_{Y|H}(y|0)$ is non-negative and $\exp(s\Lambda(y))$ is positive.

d) (4 points) Show that when $P_H(0) = P_H(1) = 1/2$, and $\mu(s) = \mu(1-s)$ for every $s \in [0,1]$, the upperbound in (b) is minimized when $s = 1/2$.

Let us recall the upper bound,i.e.,$P_H(0)^{1-s}P_H(1)^s\mu(s)$. It is easy to see that $P_H(0)^{1-s}P_H(1)^s$ is constant for all s if $P_H(0) = P_H(1)$. So we only need to show that $s = 1/2$ minimizes $\mu(s)$

Let us define $\mu'(s) = d\mu(s)/ds$. The hypothesis imply that $\mu(s) - \mu(1-s) = 0$, hence

$$\frac{d\mu(s)}{ds} - \frac{d\mu(1-s)}{ds} = \mu'(s) + \mu'(1-s) = 0.$$

Hence we have,$\mu'(s) = -\mu'(1-s)$. If we take $s = 1/2$, this leads to $\mu'(1/2) = -\mu'(1/2)$ which can only be true if $\mu'(1/2) = 0$. From (c), we have the second derivative is always non-negative, hence $\mu(s)$ is minimum in $s = 1/2$

e) (4 points) Suppose $\mu(s) = \mu(1-s)$ for every $s \in [0,1]$ but $P_H(0) > P_H(1)$. Show that the $s \in [0,1]$ that minimizes the upper bound in (b) satisfies $s > 1/2$.

Let us define $\rho = \log(P_H(1)/P_H(0))$. We have $\rho < 0$. The upper that we want to minimize is equivalent to $P_H(1)\exp(s\rho)\mu(s)$. Taking the first derivative gives us,

$$\rho P_H(0)\exp(s\rho)\mu(s) + P_H(0)\exp(s\rho)\mu'(s)$$

If we take $s = 1/2$, the second terms is equal to 0 due to our result in (d). If $\mu(1/2) = 0$, it implies that the support of $p_{Y|H}(y|0)$ and $p_{Y|H}(y|1)$ are disjoint, in this case the minimum is achieved even at $s > 1/2$. If $\mu(1/2) > 0$, then the derivative of the upper bound at $1/2$ is negative, this implies that there is $s' > 1/2$ such that the upper bound is smaller at $s = s'$ compared to $s = 1/2$.