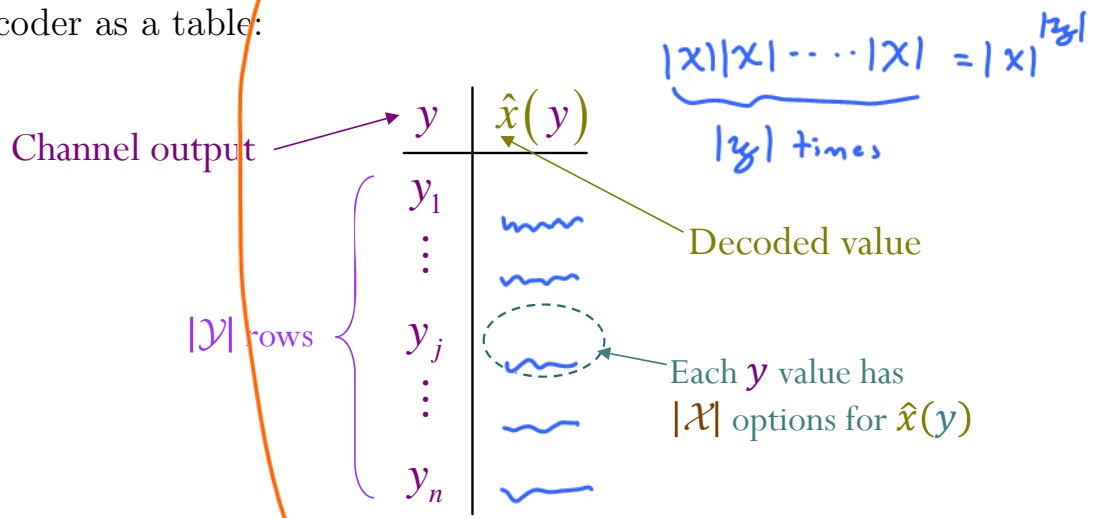


3.3 Optimal Decoding for DMC

From the previous section, we now know how to compute the error probability for any given decoder. Here, we will attempt to find the “best” decoder. Of course, by “best”, we mean “having minimum value of error probability”. It is interesting to first consider the question of how many reasonable decoders we can use.

3.29. How many “reasonable” decoders are there?: Recall from 3.21 that a decoder $\hat{x}(\cdot)$ is a function that map each observed value of the channel output y to the guessed value of the channel input. Therefore we can think of a decoder as a table:



We have already seen this table representation in Example 3.26 and Example 3.27. Such table has $|\mathcal{Y}|$ rows. For each value of y , we need to specify what is the value of $\hat{x}(y)$. To have a chance of correct guessing, any “reasonable”

decoder would select the value of $\hat{x}(y)$ from \mathcal{X} . Therefore, there are $|\mathcal{X}|^{|\mathcal{Y}|}$ reasonable decoders.

Example 3.30. The naive decoder in Example 3.25 is not a reasonable decoder. The channel input X is either 0 or 1. So, it does not make sense have a guess value of $\hat{x}(2) = 2$ or $\hat{x}(3) = 3$.

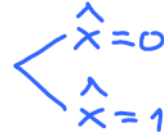


Example 3.31. “Reasonable” Decoder for BSC: For BSC in Example 3.2, any decoder has to answer two important questions:

(a) What should be the guess value of X when $Y = 0$ is observed?



(b) What should be the guess value of X when $Y = 1$ is observed?



Essentially, any reasonable decoder for the BSC needs to complete this table:

y	$\hat{x}(y)$
0	?
1	?

So, only four reasonable decoders for BSC:

y	$\hat{x}(y)$
0	0
1	1

Naive
 $\hat{x}(y) = y$

y	$\hat{x}(y)$
0	1
1	0

$\hat{x}(y) = 1 - y$
 $= \bar{y}$

y	$\hat{x}(y)$
0	0
1	0

$\hat{x}(y) \equiv 0$

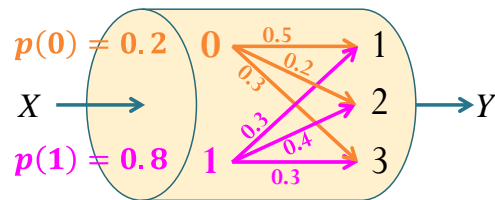
y	$\hat{x}(y)$
0	1
1	1

$\hat{x}(y) \equiv 1$

Example 3.32. For the DMC defined in Example 3.25, how many reasonable decoders are there?

We calculate the error probability of three decoders in Example 3.25, Example 3.26, and Example 3.27. There are still many reasonable possibilities

Searching for the Optimal Detector



$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 \\ 0 & \begin{bmatrix} 0.5 & 0.2 & 0.3 \end{bmatrix} \\ 1 & \begin{bmatrix} 0.3 & 0.4 & 0.3 \end{bmatrix} \end{matrix} \xrightarrow{\begin{matrix} \times 0.2 \\ \times 0.8 \end{matrix}} \begin{matrix} & 1 & 2 & 3 \\ \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} & y/x & 0 & 1 \end{matrix} = \mathbf{P}$$

y	$\hat{x}_{\text{Optimal}}(y)$
1	?
2	?
3	?



Searching for the Optimal Detector

```
% Channel Input
S_X = [0 1]; S_Y = [1 2 3];
%p_X = [0.6 0.4];
p_X = [0.2 0.8];
% Channel Characteristics
Q = [0.5 0.2 0.3; 0.3 0.4 0.3];

%% Construct all possible "reasonable" decoders
Decoder_Table_ALL = S_X';
for k = 2:length(S_Y);
    l = size(Decoder_Table_ALL,1);
    t = [];
    for i = 1:length(S_X)
        t = [t; [S_X(i)*ones(l,1) Decoder_Table_ALL]];
    end
    Decoder_Table_ALL = t;
end
```



Searching for the Optimal Detector

```
%% Calculate the error probability for each of the decoder
PE_ALL = [];
for k = 1:size(Decoder_Table_ALL,1)
    Decoder_Table = Decoder_Table_ALL(k,:);
    PC = 0;
    for k = 1:length(S_X)
        I = (Decoder_Table == S_X(k));
        Q_row = Q(k,:);
        PC = PC + p_X(k)*sum(Q_row(I));
    end
    PE_theretical = 1-PC;
    PE_ALL = [PE_ALL; PE_theretical];
end

%% Display the results
[Decoder_Table_ALL PE_ALL]

%% Find the optimal detectors
Min_PE = min(PE_ALL)
I = [abs(PE_ALL-Min_PE)<1e-10];
Optimal_Detector = Decoder_Table_ALL(I,:)
```



Searching for the Optimal Detector

```
>> DMC_decoder_ALL_demo
```

```
ans =    $\hat{x}(1)$        $\hat{x}(2)$        $\hat{x}(3)$        $P(\mathcal{E})$ 
        0          0          0      0.8000
        0          0          1      0.6200
        0          1          0      0.5200
        0          1          1      0.3400
        1          0          0      0.6600
        1          0          1      0.4800
        1          1          0      0.3800
        1          1          1      0.2000
```

Ex. 3.26

Ex. 3.27

```
Min_PE =
```

```
0.2000
```

```
Optimal_Detector =
```

```
1      1      1
```



to evaluate. Using MATLAB, we can find the error probability for all possible reasonable decoders:

y	$\hat{x}_{D1}(y)$	$\hat{x}_{D2}(y)$	$\hat{x}_{D3}(y)$	$\hat{x}_{D4}(y)$	$\hat{x}_{D5}(y)$	$\hat{x}_{D6}(y)$	$\hat{x}_{D7}(y)$	$\hat{x}_{D8}(y)$
1	0	0	0	0	1	1	1	1
2	0	0	1	1	0	0	1	1
3	0	1	0	1	0	1	0	1
$P(\mathcal{E})$	0.80	0.62	0.52	0.34	0.66	0.48	0.38	0.20

3.33. For general DMC, it would be tedious to list all possible decoders. It is even more time-consuming to try to calculate the error probability for all of them. Therefore, in this section, we will derive a visual construction and a formula of the “optimal” decoder.

3.34. From the recipe 3.28 for finding $P(\mathcal{C})$ and $P(\mathcal{E})$, we see that $P(\mathcal{C})$ is the sum of our circled numbers. So, to maximize $P(\mathcal{C})$, we want to circle the largest number. For row y in the decoding table, whatever the value we select for $\hat{x}(y)$ will determine which number will be circled in the column corresponding to y in **matrix P**. To maximize $P(\mathcal{C})$, we want to circle the largest number in the column. This means $\hat{x}(y)$ should be the same as the x value that maximizes the probability value in the corresponding column.

Example 3.35. Find the optimal decode and the corresponding error probability for the DMC and the input probabilities defined in Example 3.17.

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 \\ 0 & \begin{bmatrix} 0.5 & 0.2 & 0.3 \end{bmatrix} \\ 1 & \begin{bmatrix} 0.3 & 0.4 & 0.3 \end{bmatrix} \end{matrix} \xrightarrow{\begin{matrix} \times 0.2 \\ \times 0.8 \end{matrix}} \begin{matrix} \hat{x}(y) \\ 1 & 2 & 3 \\ \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} \\ 0 \\ 1 \end{matrix} = \mathbf{P}$$

$$P(\mathcal{C}) = 0.24 + 0.32 + 0.24 = 0.4$$

$$P(\mathcal{E}) = 1 - 0.4 = 0.6$$

y	$\hat{x}_{\text{Optimal}}(y)$
1	1
2	1
3	1

3.36. Deriving the optimal decoder: Mathematically, we first note that to minimize $P(\mathcal{E})$, we need to maximize $P(\mathcal{C})$. Here, we apply the total

probability theorem by using the events $[Y = y]$ to partition the sample space:

$$P(\mathcal{C}) = \sum_y P(\mathcal{C} | [Y = y]) P[Y = y].$$

Event \mathcal{C} is the event $[\hat{X} = X]$. Therefore,

$$P(\mathcal{C}) = \sum_y P[\hat{X} = X | Y = y] P[Y = y].$$

Now, recall that our decoder is a function of Y ; that is $\hat{X} = \hat{x}(Y)$. So,

$$\begin{aligned} P(\mathcal{C}) &= \sum_y P[\hat{x}(Y) = X | Y = y] P[Y = y] \\ &= \sum_y P[X = \hat{x}(y) | Y = y] P[Y = y] \end{aligned}$$

In this form, we see¹⁴ that for each $Y = y$, we should maximize $P[X = \hat{x}(y) | Y = y]$. Therefore, for each y , the decoder $\hat{x}(y)$ should output the value of x which maximizes¹⁵ $P[X = x | Y = y]$:

$$\hat{x}_{\text{optimal}}(y) = \arg \max_x P[X = x | Y = y].$$

In other words, the *optimal* decoder is the decoder that maximizes the “a posteriori probability” $P[X = x | Y = y]$.

Definition 3.37. The optimal decoder derived in 3.36 is called the **maximum a posteriori probability (MAP) decoder**:

$$\hat{x}_{\text{MAP}}(y) = \hat{x}_{\text{optimal}}(y) = \arg \max_x P[X = x | Y = y]. \quad (8)$$

3.38. After the fact, it is quite intuitive that this should be the best decoder.

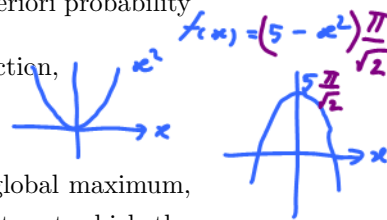
Recall that the decoder don’t have a direct access to the X value.

¹⁴We also see that any decoder that produces random results (on the support of X) can not be better than our optimal decoder. Outputting the value of x which does not maximize the a posteriori probability reduces the contribution in the sum that gives $P(\mathcal{C})$.

¹⁵For those who are not familiar with the “arg max” (arguments of the maximum) function,

$$\arg \max_x f(x) = \text{the } x \text{ value that maximizes } f(x).$$

The corresponding maximum value of $f(x)$ is $\max_x f(x)$. In other words, in contrast to global maximum, referring to the largest outputs of a function, $\arg \max$ refers to the inputs, or arguments, at which the function outputs are as large as possible. For example, for $f(x) = 5 - x^2$, we have $\arg \max_x f(x) = 0$ and $\max_x f(x) = 5$.



$$\begin{aligned} \max_x f(x) &= 5 \frac{\pi}{12} \\ \arg \max_x f(x) &= 0 \end{aligned}$$

- Without knowing the value of Y , to minimize the error probability, it should guess the most likely value of X which is the value of x that maximize $P[X = x]$.
- Knowing $Y = y$, the decoder can update its probability about x from $P[X = x]$ to $P[X = x|Y = y]$. Therefore, the decoder should guess the value of the most likely x value conditioned on the fact that $Y = y$.

3.39. We should manipulate Formula (3.37) for the MAP decoder a bit further because, in practice, we usually only know $p(x)$ and $Q(y|x)$. To connect these terms to $P[X = x|Y = y]$ required in (3.37), first, recall “Form 1” of the Bayes’ theorem:

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)}.$$

Here, we set $B = [X = x]$ and $A = [Y = y]$.

$$P[X = x | Y = y] = \frac{P[Y = y | X = x] P[X = x]}{P[Y = y]}$$

Therefore,

$$\hat{x}_{\text{MAP}}(y) = \arg \max_x Q(y|x)p(x). \quad (9)$$

Handwritten notes:
 - A purple circle highlights $Q(y|x)$ and a green circle highlights $p(x)$ in equation (9).
 - A blue arrow points from $P[Y=y]$ in the fraction above to the denominator of (9).
 - A green arrow points from $P[X=x]$ in the fraction above to $p(x)$ in (9).
 - A blue arrow points from $P[Y=y]$ to the text: "Does not depend on x and it is '>0' => can be ignored".

Note that the term $P[Y = y]$ does not depend on x and it is positive; therefore, it does not change the the answer of $\arg \max$ and hence can be ignored.

3.40. A recipe for finding the MAP decoder (optimal decoder) and its corresponding error probability:

- Find the **P matrix** by scaling elements in each row of the **Q** matrix by their corresponding prior probability $p(x)$.
 - Select (by circling) the maximum value in each column (for each value of y) in the **P** matrix.
 - If there are multiple max values in a column, select one. This won't affect the optimality of your answer.
- (i) The corresponding x value is the value of \hat{x} for that y .

Review: EES315 (2020)

6.4. Interpretation: It is sometimes useful to interpret $P(A)$ as our **knowledge** of the occurrence of event A **before** the **experiment takes place**. Conditional probability²⁴ $P(A|B)$ is the **updated probability** of the event A given that we now know that B occurred (but we still do not know which particular outcome in the set B did occur). *In general, $P(A)$ and $P(A|B)$ are not the same.*

Definition 6.5. Sometimes, we refer to $P(A)$ as

- a **priori** probability, or
- the **prior probability** of A , or
- the **unconditional probability** of A .

$P(A|B)$

| a posteriori probability

| the posterior probability

| conditional probability



(ii) The **sum** of the selected values from the **P** matrix is **$P(\mathcal{C})$** .

(c) **$P(\mathcal{E}) = 1 - P(\mathcal{C})$** .

Example 3.41. We have applied recipe 3.40 back when we try to find the optimal decoder in Example 3.35.

Example 3.42. Find the MAP decoder and its corresponding error probability for the DMC channel whose **Q** matrix is given by

$$\mathbf{Q} = \begin{array}{c|ccc} x \setminus y & 1 & 2 & 3 \\ \hline 0 & 0.5 & 0.2 & 0.3 \\ 1 & 0.3 & 0.4 & 0.3 \end{array} \xrightarrow{\substack{\times 0.6 \\ \times 0.4}} \begin{array}{ccc|c} & 1 & 2 & 3 & y/x \\ \hline 0 & 0.30 & 0.12 & 0.18 & 0 \\ 1 & 0.12 & 0.16 & 0.12 & 1 \end{array} = \mathbf{P}$$

y	$\hat{x}_{\text{MAP}}(y)$
1	0
2	1
3	0

and **$\mathbf{p} = [0.6, 0.4]$** . Note that the DMC is the same as in Example 3.25 but the input probabilities are different.

$$\begin{aligned} P(\mathcal{C}) &= 0.30 + 0.18 + 0.16 = 0.64 \\ P(\mathcal{E}) &= 1 - 0.64 = 0.36 \end{aligned}$$

Definition 3.43. In many scenarios, the **MAP** decoder is **too complicated** or the **prior probabilities** are unknown. In such cases, we may consider using a **suboptimal** decoder that ignores the prior probability term in (9). This decoder is called the **maximum likelihood (ML)** decoder:

$$\hat{x}_{\text{ML}}(y) = \arg \max_x Q(y|x). \quad (10)$$

3.44. **ML decoder** is the **same as the MAP decoder** when **X is a uniform random variable**. In other words, when the prior probabilities $p(x)$ are uniform, the ML decoder is **optimal**.

3.45. A **recipe** for finding the ML decoder and its corresponding error probability:

(a) **Select** (by circling) the **maximum value in each column** (for each value of y) in the **Q** matrix.

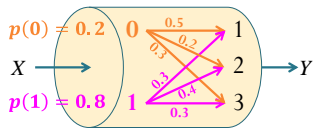
- If there are **multiple max** values in a column, **select one**.

Different choices will lead to different $P(\mathcal{E})$. However, if the information about **p** is not available at the decoder, it can not determine which choice is better anyway.

- The corresponding x value is the value of \hat{x} for that y .
- Find the \mathbf{P} matrix by scaling elements in each row of the \mathbf{Q} matrix by their corresponding prior probability $p(x)$.
 - In the \mathbf{P} matrix, select the elements corresponding to the selected positions in the \mathbf{Q} matrix.
 - The sum of the selected values from the \mathbf{P} matrix is $P(\mathcal{C})$.
 - $P(\mathcal{E}) = 1 - P(\mathcal{C})$.

Example 3.46. Find an ML decoder and its corresponding error probability for the DMC channel and the input probabilities defined in Example 3.17.

$$P = [0.2 \ 0.8]$$



Sol. ✳ 1

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \end{matrix} \xrightarrow{\begin{matrix} \times 0.2 \\ \times 0.8 \end{matrix}} \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} \end{matrix} = \mathbf{P}$$

y	$\hat{x}_{ML}(y)$
1	0
2	1
3	0

$$P(\mathcal{C}) = 0.10 + 0.32 + 0.06 = 0.48$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 0.52$$

Sol. ✳ 2

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \end{matrix} \xrightarrow{\begin{matrix} \times 0.2 \\ \times 0.8 \end{matrix}} \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} \end{matrix} = \mathbf{P}$$

y	$\hat{x}_{ML}(y)$
1	0
2	1
3	1

$$P(\mathcal{C}) = 0.10 + 0.32 + 0.24 = 0.66$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 0.34$$

Note that, for ML decoders, when there are multiple solutions, they may not give the same $P(\mathcal{E})$.

Different Decoders

• For naïve decoder, $\hat{x}_{\text{naïve}}(y) = y$. Therefore, we can simply copy the values in the y -column into its decoding table.

• For DIY decoder, the decoding table is provided. (Alternatively, some equation may be given; in which case, we can simply plug-in each of the possible y values to get $\hat{x}_{\text{DIY}}(y)$.)

y	$\hat{x}_{\text{naïve}}(y)$	$\hat{x}_{\text{DIY}}(y)$	$\hat{x}_{\text{MAP}}(y)$	$\hat{x}_{\text{ML}}(y)$
1	1	1	1	0
2	2	1	1	1
3	3	0	1	0

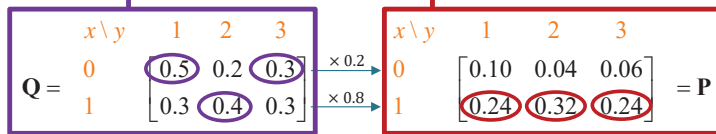
[Ex. 3.25]

[Ex. 3.27]

[Ex. 3.35]

[Ex. 3.46]

• Deriving the MAP and ML decoder follows almost the same recipe: select the max value in each column and read the corresponding x -value. The difference is that the MAP decoder uses the \mathbf{P} matrix but the ML decoder uses the \mathbf{Q} matrix.



Finding $P(\mathcal{E})$

- Once a decoder $\hat{x}(y)$ is defined, we can find its corresponding $P(\mathcal{E})$ easily from the \mathbf{P} matrix:
 - Write $\hat{x}(y)$ values on top of the y values for the \mathbf{P} matrix.
 - For each column y in the \mathbf{P} matrix, circle the element whose corresponding x value is the same as $\hat{x}(y)$.

y	$\hat{x}_{\text{naïve}}(y)$	$\hat{x}_{\text{DIY}}(y)$	$\hat{x}_{\text{MAP}}(y)$	$\hat{x}_{\text{ML}}(y)$
1	1	1	1	0
2	2	1	1	1
3	3	0	1	0
$P(\mathcal{E})$	0.76	0.38	0.20	0.52

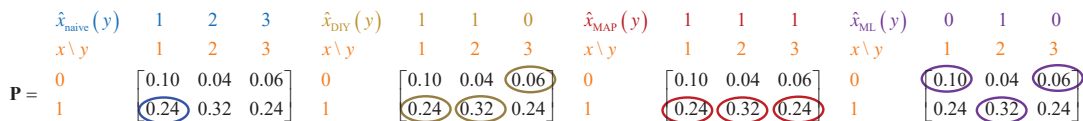
[Ex. 3.25]

[Ex. 3.27]

[Ex. 3.35]

[Ex. 3.46]

- $P(\mathcal{C})$ = the sum of the circled probabilities.
- $P(\mathcal{E}) = 1 - P(\mathcal{C})$.



ECS 452: In-Class Exercise # 8

Instructions

1. Separate into groups of no more than three students each. **The group cannot be the same as any of your former groups.** Only one submission is needed for each group.
2. [ENRE] Explanation is not required for this exercise.
3. [WASP] Write your answer(s) in the corresponding space(s) provided.
4. Do not panic.

Date: 14 / 2 / 2020			
Name			ID <small>(last 3 digits)</small>

1. Consider a DMC whose transition matrix \mathbf{Q} and joint pmf matrix \mathbf{P} are given below.

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0.2 & 0.2 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.2 & 0.3 \\ 0.3 & 0.1 & 0.1 & 0.5 \\ 0.1 & 0.1 & 0.7 & 0.1 \end{bmatrix} \end{matrix}$$

$$\mathbf{P} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0.08 & 0.08 & 0.12 & 0.12 \\ 0.03 & 0.12 & 0.06 & 0.09 \\ 0.06 & 0.02 & 0.02 & 0.10 \\ 0.01 & 0.01 & 0.07 & 0.01 \end{bmatrix} \end{matrix}$$

We will consider four decoders:

- (a) **The naïve decoder.** For naïve decoder, $\hat{x}_{\text{naïve}}(y) = y$. Therefore, we can simply copy the values in the y -column into its decoding table.
- (b) **A DIY decoder defined by $\hat{x}_{\text{DIY}}(y) = 5 - y$.** The equation for the decoder is given; so we can simply plug-in each of the possible y values to get $\hat{x}_{\text{DIY}}(y)$.
- (c) **The MAP decoder.**
- (d) **The ML decoder.**

Deriving the MAP and ML decoder follows almost the same recipe: select the max value in each column and read the corresponding x -value. The difference is that the **MAP decoder uses the \mathbf{P} matrix** but the **ML decoder uses the \mathbf{Q} matrix**.

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0.2 & 0.2 & 0.3 & 0.3 \\ 0.1 & 0.4 & 0.2 & 0.3 \\ 0.3 & 0.1 & 0.1 & 0.5 \\ 0.1 & 0.1 & 0.7 & 0.1 \end{bmatrix} \end{matrix}$$

$$\mathbf{P} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0.08 & 0.08 & 0.12 & 0.12 \\ 0.03 & 0.12 & 0.06 & 0.09 \\ 0.06 & 0.02 & 0.02 & 0.10 \\ 0.01 & 0.01 & 0.07 & 0.01 \end{bmatrix} \end{matrix}$$

Find the decoding table for each of these decoders. Also, find their probabilities of decoding error. Put your answers in the table below.

y	$\hat{x}_{\text{naïve}}(y)$	$\hat{x}_{\text{DIY}}(y)$	$\hat{x}_{\text{MAP}}(y)$	$\hat{x}_{\text{ML}}(y)$
1	1	4	1	3
2	2	3	2	2
3	3	2	1	4
4	4	1	1	3
$P(\mathcal{E})$	0.77	0.79	0.56	0.65

Once a decoder $\hat{x}(y)$ is defined, we can find its corresponding $P(\mathcal{E})$ easily from the \mathbf{P} matrix:

For each column y in the \mathbf{P} matrix, circle the element whose corresponding x value is the same as $\hat{x}(y)$; then, $P(\mathcal{C}) =$ the sum of the circled probabilities; and $P(\mathcal{E}) = 1 - P(\mathcal{C})$.

	$\hat{x}_{\text{naïve}}(y)$	$\hat{x}_{\text{DIY}}(y)$	$\hat{x}_{\text{MAP}}(y)$
	$\hat{x}_{\text{ML}}(y)$		
	\mathbf{P}		

EES 452: Exercise # 9

Instructions

1. Work in your assigned group.
2. Do not use red ink.
3. Only one submission is needed for each group.
4. Submit your work via the corresponding exercise posted in Google Classroom.
5. **Do not panic.**

Date: 23 / 2 / 2021			
Name			ID <small>(last 3 digits)</small>

1. Consider a DMC whose transition matrix \mathbf{Q} and joint probability matrix \mathbf{P} are given below.

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.4 & 0.1 & 0.1 & 0.4 \\ 2 & 0.2 & 0.1 & 0.5 & 0.2 \\ 3 & 0.1 & 0.7 & 0.1 & 0.1 \\ 4 & 0.2 & 0.2 & 0.2 & 0.4 \end{matrix}$$

$$\mathbf{P} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.08 & 0.02 & 0.02 & 0.08 \\ 2 & 0.10 & 0.05 & 0.25 & 0.10 \\ 3 & 0.02 & 0.14 & 0.02 & 0.02 \\ 4 & 0.02 & 0.02 & 0.02 & 0.04 \end{matrix} \begin{matrix} \xrightarrow{\Sigma} 0.2 \\ \xrightarrow{\Sigma} 0.5 \\ \xrightarrow{\Sigma} 0.2 \\ \xrightarrow{\Sigma} 0.1 \end{matrix}$$

(a) Find the input probability vector \mathbf{p} .

The input probability can be found from the joint probability matrix \mathbf{P} by summing along each of the rows. This gives $\mathbf{p} = [0.2 \ 0.5 \ 0.2 \ 0.1]$.
 Alternatively, recall that the elements in \mathbf{P} can be found from \mathbf{Q} by scaling each row of \mathbf{Q} by $p(x)$. Therefore, to get the value of $p(x)$, we simply have the elements in \mathbf{P} divided by the corresponding elements in \mathbf{Q} . For example, $p(1)$ can be calculated from $\frac{0.08}{0.4}$ or $\frac{0.02}{0.1}$ or $\frac{0.08}{0.4}$.

(b) Consider three decoders:

- i. A DIY decoder defined by $\hat{x}_{\text{DIY}}(y) = 2 + (-1)^y$.
- ii. The MAP decoder.
- iii. The ML decoder.

The equation for the decoder is given; so we can simply plug-in each of the possible y values to get $\hat{x}_{\text{DIY}}(y)$.

Deriving the MAP and ML decoder follows almost the same recipe: select the max value in each column and read the corresponding x -value. The difference is that the MAP decoder uses the \mathbf{P} matrix but the ML decoder uses the \mathbf{Q} matrix.

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.4 & 0.1 & 0.1 & 0.4 \\ 2 & 0.2 & 0.1 & 0.5 & 0.2 \\ 3 & 0.1 & 0.7 & 0.1 & 0.1 \\ 4 & 0.2 & 0.2 & 0.2 & 0.4 \end{matrix}$$

For the last column, there are two max. Selecting any of these are OK for the ML decoder. Note that the corresponding $P(\mathcal{E})$ won't be the same.

$$\mathbf{P} = \begin{matrix} x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.08 & 0.02 & 0.02 & 0.08 \\ 2 & 0.10 & 0.05 & 0.25 & 0.10 \\ 3 & 0.02 & 0.14 & 0.02 & 0.02 \\ 4 & 0.02 & 0.02 & 0.02 & 0.04 \end{matrix}$$

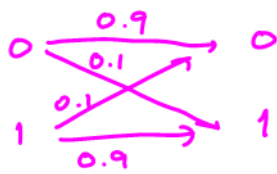
Find the decoding table for each of these decoders. Also, find their probabilities of decoding error. Put your answers in the table below.

Once a decoder $\hat{x}(y)$ is defined, we can find its corresponding $P(\mathcal{E})$ easily from the \mathbf{P} matrix:
 For each column y in the \mathbf{P} matrix, circle the element whose corresponding x value is the same as $\hat{x}(y)$; then, $P(\mathcal{C})$ = the sum of the circled probabilities; and $P(\mathcal{E}) = 1 - P(\mathcal{C})$.

y	$\hat{x}_{\text{DIY}}(y)$	$\hat{x}_{\text{MAP}}(y)$	$\hat{x}_{\text{ML}}(y)$
1	1	2	1
2	3	3	3
3	1	2	2
4	3	2	1 or 4
$P(\mathcal{E})$	0.74	0.41	0.45 or 0.49
$P(\mathcal{C})$	0.26	0.59	0.55 or 0.51

$$\hat{x}_{\text{DIY}}(y) \begin{matrix} 1 & 3 & 1 & 3 \\ x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.08 & 0.02 & 0.02 & 0.08 \\ 2 & 0.10 & 0.05 & 0.25 & 0.10 \\ 3 & 0.02 & 0.14 & 0.02 & 0.02 \\ 4 & 0.02 & 0.02 & 0.02 & 0.04 \end{matrix} \quad \hat{x}_{\text{MAP}}(y) \begin{matrix} 2 & 3 & 2 & 2 \\ x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.08 & 0.02 & 0.02 & 0.08 \\ 2 & 0.10 & 0.05 & 0.25 & 0.10 \\ 3 & 0.02 & 0.14 & 0.02 & 0.02 \\ 4 & 0.02 & 0.02 & 0.02 & 0.04 \end{matrix}$$

$$\hat{x}_{\text{ML}}(y) \begin{matrix} 1 & 3 & 2 & 1 \text{ or } 4 \\ x \backslash y & 1 & 2 & 3 & 4 \\ 1 & 0.08 & 0.02 & 0.02 & 0.08 \\ 2 & 0.10 & 0.05 & 0.25 & 0.10 \\ 3 & 0.02 & 0.14 & 0.02 & 0.02 \\ 4 & 0.02 & 0.02 & 0.02 & 0.04 \end{matrix}$$



y	$\hat{x}_{MAP}(y)$	
0	0	$p(1) = 1 - p(0) = 1 - 0.8 = 0.2$
1	1	$p(0) \equiv P[X=0]$

$p(x) \equiv P[X=x]$

Example 3.47. Find the ML decoder and the corresponding error probability for a communication over BSC with $p = 0.1$ and $p(0) = 0.8$.

$Q = \begin{matrix} x \setminus y & 0 & 1 \\ 0 & 1-p & p \\ 1 & p & 1-p \end{matrix} = \begin{matrix} \hat{x}_{ML}(y) \setminus y & 0 & 1 \\ 0 & 0.9 & 0.1 \\ 1 & 0.1 & 0.9 \end{matrix} \xrightarrow{\begin{matrix} \times 0.8 \\ \times 0.2 \end{matrix}} \begin{matrix} x \setminus y & 0 & 1 \\ 0 & 0.72 & 0.08 \\ 1 & 0.02 & 0.18 \end{matrix} = P$

$P(C) = 0.72 + 0.18 = 0.90$
 $P(E) = 1 - 0.90 = 0.10$

Note that

- the prior probabilities $p(0)$ (and $p(1)$) is not used when finding \hat{x}_{ML} ,
- the ML decoder and the MAP decoder are the same in this example.
 - ML decoder can be optimal even when the prior probabilities are not uniform.

3.48. MAP vs. ML:

$$\begin{aligned}
 \hat{x}_{MAP}(y) &= \arg \max_x p_{X,Y}(x, y) \\
 &= \hat{x}_{\text{optimal}}(y) \quad \text{a posteriori probability} \\
 &= \arg \max_x P[X=x|Y=y] \\
 &= \arg \max_x Q(y|x) \overbrace{p(x)}^{\text{prior probability}}
 \end{aligned}$$

- Decoder is derived from the **P** matrix
- Select (by circling) the maximum value in each column (for each value of y) in the **P** matrix.
- The corresponding x value is the value of $\hat{x}_{MAP}(y)$.

$$\hat{x}_{ML}(y) = \arg \max_x \overbrace{Q(y|x)}^{\text{likelihood function}}$$

Optimal at least when $p(x)$ is uniform (the channel inputs are equally likely)

Can be derived without knowing the channel input probabilities.

- Decoder is derived from the **Q** matrix
- Select (by circling) the maximum value in each column (for each value of y) in the **Q** matrix.
- The corresponding x value is the value of $\hat{x}_{ML}(y)$.

- Once the decoder (the decoding table) is derived $P(C)$ and $P(E)$ are calculated by adding the corresponding probabilities in the **P** matrix.

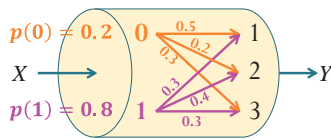
MAP Decoder: Two Approaches

[3.37-3.39]

$$\begin{aligned}
 \hat{x}_{\text{MAP}}(y) &= \arg \max_x p_{X,Y}(x, y) \\
 &= \hat{x}_{\text{optimal}}(y) \quad \text{a posteriori probability} \\
 &= \arg \max_x P[X = x | Y = y] \\
 &= \arg \max_x Q(y|x) p(x) \quad \text{prior probability}
 \end{aligned}$$

[3.40]

- 1) Decoder is derived from the **P** matrix
- 2) Select (by circling) the maximum value in each column (for each value of y) in the **P** matrix.
- 3) The corresponding x value is the value of $\hat{x}_{\text{MAP}}(y)$.



$$\begin{array}{c}
 \begin{matrix} 1 & 1 & 1 & \hat{x}_{\text{MAP}}(y) \\ 1 & 2 & 3 & \\ \begin{matrix} 1 \\ 0 \end{matrix} \backslash y & \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} & \xrightarrow{\begin{matrix} \times 0.2 \\ \times 0.8 \end{matrix}} & \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} & \begin{matrix} y/x \\ 0 \\ 1 \end{matrix} \\
 \mathbf{Q} = & & & & = \mathbf{P}
 \end{array}$$