

## 3.2 Decoder and Symbol Error Probability

**3.19.** Knowing the characteristics of the channel and its input, on the receiver side, we can use this information to build a “good” receiver.

We now consider a part of the receiver called the **(channel) decoder**. Its job is to **guess the value of the channel input**<sup>12</sup>  $X$  from the value of the received channel output  $Y$ . We denote this **guessed value** by  $\hat{X}$ .

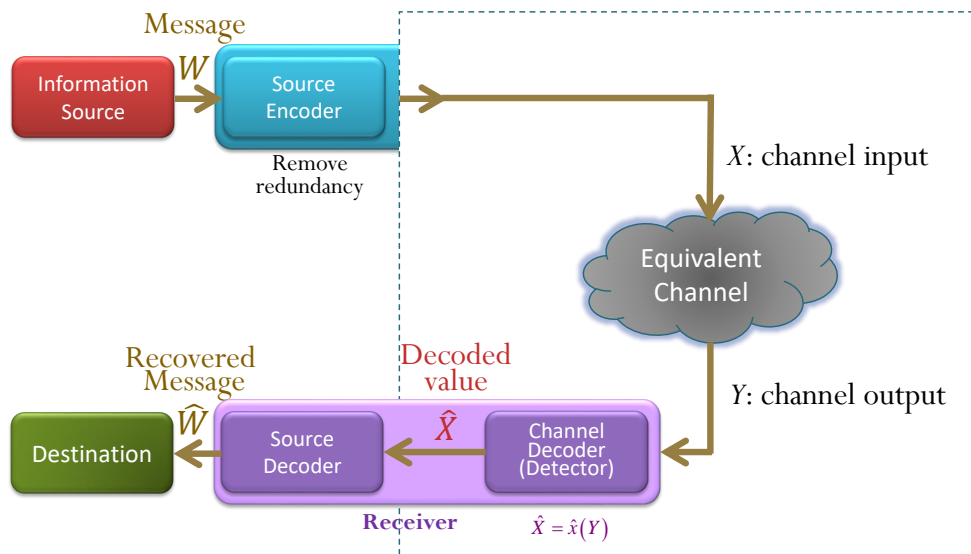


Figure 11: Adding a (channel) decoder to improve the performance of the system in Figure 4 considered in Section 3.1.

**3.20.** A “good” decoder is the one that (often) **guesses correctly**. So, our goal here is to

- maximize the probability of correct guessing  $\rightarrow P(C)$
- minimize the probability of wrong guessing  $\rightarrow P(E)$

Quantitatively, to measure the performance of a decoder, we define a quantity called the (symbol) error probability.

<sup>12</sup>To simplify the analysis, we still haven’t considered the channel encoder. (It may be there but is included in the equivalent channel or it may not be in the system at all.)

**Definition 3.21.** The (symbol) **error probability**, denoted by  $P(\mathcal{E})$ , can be calculated from

$$P(\mathcal{E}) = P[\hat{X} \neq X].$$

- $\mathcal{E}$  is the (decoding) error event.
- A dual (complementary) quantity is the probability of correct decoding:

$$P(\mathcal{C}) = P[\hat{X} = X] = 1 - P(\mathcal{E}).$$

- For channel with binary input, the error probability is the same as bit error rate (BER).

**3.22.** A “reasonable” decoder should make a guess based on all the information it has obtained. Here, the only information it can observe is the value of  $Y$ . Therefore, a decoder is a function of  $Y$ , say,  $g(Y)$ . Therefore,  $\hat{X} = g(Y)$ .

We use  $\hat{x}(\cdot)$  to denote this function  $g(\cdot)$ ; So,  $\hat{X} = \hat{x}(Y)$ . It is a deterministic function operating on the random channel output  $Y$ ; the randomness in the decoded value  $\hat{X}$  comes from the randomness in  $Y$ .

**Definition 3.23.** A “naive” decoder is a decoder that simply sets  $\hat{X} = Y$ . Equivalently,

$$\hat{x}_{\text{naive}}(y) = y.$$

- We may call this a “pass-through” decoder or a “do-nothing” decoder.
- It usually has poor performance. However, we shall see that, in some cases, this simple decoder can be optimal<sup>13</sup>.

**3.24.** For general DMC, the error probability of the naive decoder is

$$\begin{aligned} P(\mathcal{E}) &= P[\hat{X} \neq X] = P[Y \neq X] = 1 - P[Y = X] \\ &= 1 - \sum_x P[Y = x, X = x] = 1 - \sum_x P[Y = x | X = x]P[X = x] \\ &= 1 - \sum_x Q(x|x)p(x) \end{aligned}$$

---

<sup>13</sup>“Optimal” means nothing has better error probability in that particular situation.

**Example 3.25.** Consider the BAC channel and input probabilities specified in Example 3.16. Find  $P(\mathcal{E})$  when  $\hat{X} = Y$

(a) Method 1:

$$\mathbf{Q} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \xrightarrow{\substack{\times 0.5 \\ \times 0.5}} \begin{bmatrix} 0.35 & 0.15 \\ 0.20 & 0.30 \end{bmatrix} = \mathbf{P}$$

$$P(\mathcal{C}) \equiv P[\hat{X} = X] = P[Y = X] = 0.35 + 0.30 = 0.65$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 1 - 0.65 = 0.35$$

(b) Method 2: To calculate the error probability directly, we write

$$\begin{aligned}
 P(\mathcal{E}) &= P[\hat{X} \neq X] = P[Y \neq X] \\
 &= P[Y = 0, X = 1] + P[Y = 1, X = 0] \\
 &= P[Y = 0 | X = 1] P[X = 1] + P[Y = 1 | X = 0] P[X = 0] \\
 &= 0.4 \times \frac{1}{2} + 0.3 \times \frac{1}{2} = 0.35.
 \end{aligned}$$

(c) Method 3: Using the derived formula from 3.24,

$$\begin{aligned}
 P(\mathcal{E}) &= 1 - (Q(0|0)p(0) + Q(1|1)p(1)) \\
 &= 1 - \left(0.7 \times \frac{1}{2} + 0.6 \times \frac{1}{2}\right) = 0.35.
 \end{aligned}$$

**Example 3.26.** Continue from Examples 3.12 and 3.18. Find the error probability  $P(\mathcal{E})$  when a naive decoder is used with a DMC channel in which  $\mathcal{X} = \{0, 1\}$ ,  $\mathcal{Y} = \{1, 2, 3\}$ ,  $\mathbf{Q} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$  and  $\underline{\mathbf{p}} = [0.2, 0.8]$ .

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

$$P(\mathcal{C}) = P[\hat{X} = X] = P[Y = X] = 0.24$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 1 - 0.24 = 0.76$$

**Example 3.27. DIY Decoder:** Consider a different decoder specified in the **decoding table** below. Find the error probability  $P(\mathcal{E})$  when such decoder is used in Example 3.26.

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

$$P = \begin{matrix} & \hat{x} & 0 & 1 & 0 \\ \begin{matrix} \swarrow y \\ 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} \end{matrix}$$

$$P(\mathcal{C}) = P[\hat{X} = X] = 0.10 + 0.32 + 0.06 = 0.48$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 1 - 0.48 = 0.52$$

**Example 3.28.** Repeat Example 3.27 but use the following decoder

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

$$P = \begin{matrix} & \hat{x} & 1 & 1 & 0 \\ \begin{matrix} \swarrow y \\ 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.10 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} \end{matrix}$$

$$P(\mathcal{C}) = P[\hat{X} = X] = 0.24 + 0.32 + 0.06 = 0.62$$

$$P(\mathcal{E}) = 1 - P(\mathcal{C}) = 1 - 0.62 = 0.38$$

Observation: For each column of the  $\mathbf{P}$  matrix, we circle the probability corresponding to the row of  $x$  that has the same value as  $\hat{x}(y)$ .

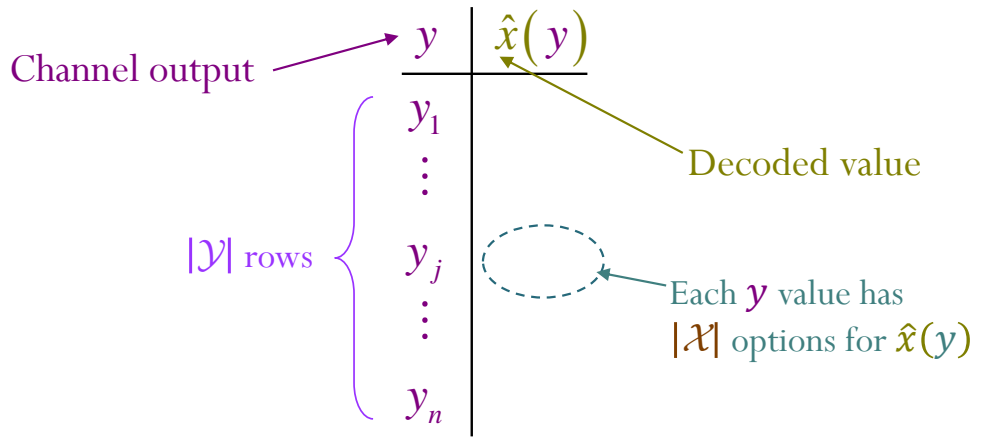
**3.29.** A **recipe** for finding  $P(\mathcal{E})$  of any (DIY) decoder:

- Find the  $\mathbf{P}$  matrix by scaling each row of the  $\mathbf{Q}$  matrix by its corresponding  $p(x)$ .
- Write  $\hat{x}(y)$  values on top of the  $y$  values for the  $\mathbf{P}$  matrix.
- For **column**  $y$  in the  $\mathbf{P}$  matrix, **circle the element** whose corresponding  $x$  value is the same as  $\hat{x}(y)$ .
- $P(\mathcal{C})$  = the **sum of the circled probabilities**.  $P(\mathcal{E}) = 1 - P(\mathcal{C})$ .

**3.3 Optimal Decoding for DMC**

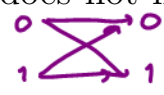
From the previous section, we now know how to compute the error probability for any given decoder. Here, we will attempt to find the “best” decoder. Of course, by “best”, we mean “having minimum value of error probability”. It is interesting to first consider the question of how many reasonable decoders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

**3.30.** How many “reasonable” decoders are there?: Recall from 3.22 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:



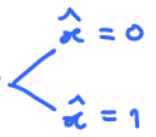
We have already seen this table representation in Example 3.27 and Example 3.28. Such table has  $|\mathcal{Y}|$  rows. For each value of  $y$ , we need to specify what is the value of  $\hat{x}(y)$ . To have a chance of correct guessing, any “reasonable” decoder would select the value of  $\hat{x}(y)$  from  $\mathcal{X}$ . Therefore, there are  $|\mathcal{X}|^{|\mathcal{Y}|}$  reasonable decoders.

**Example 3.31.** The naive decoder in Example 3.26 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  $\hat{x}(3) = 3$ .

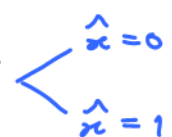


**Example 3.32.** “Reasonable” Decoder for BSC: For BSC in Example 3.2, any decoder has to answer two important questions:

(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:

$y \mid \hat{x}(y)$	$y \mid \hat{x}(y)$	$y \mid \hat{x}(y)$	$y \mid \hat{x}(y)$
0 $\mid$ 0	0 $\mid$ 1	0 $\mid$ 0	0 $\mid$ 1
1 $\mid$ 1	1 $\mid$ 0	1 $\mid$ 0	1 $\mid$ 1
Naive			
$\hat{x}(y) = y$	$\hat{x}(y) = \bar{y}$ $= 1 - y$	$\hat{x}(y) \equiv 0$	$\hat{x}(y) \equiv 1$

**Example 3.33.** For the DMC defined in Example 3.26, how many reasonable decoders are there?

$$|\mathcal{X}|^{|\mathcal{Y}|} = 2^3 = 8$$

We calculate the error probability of three decoders in Example 3.26, Example 3.27, and Example 3.28. There are still many reasonable possibilities 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

Ex 3.27

Ex 3.28

**3.34.** 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.

**3.35.** From the recipe 3.29 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  $\mathbf{P}$ . 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 3.36.** For the DMC and the input probabilities defined in Example 3.26, the joint pmf matrix  $\mathbf{P}$  was found to be

$$\mathbf{P} = \begin{array}{c|ccc}
 \hat{x} & \underline{1} & \underline{1} & \underline{1} \\
 x \setminus y & 1 & 2 & 3 \\
 0 & 0.1 & 0.04 & 0.06 \\
 1 & \underline{0.24} & \underline{0.32} & \underline{0.24}
 \end{array}$$

$P(\mathcal{C}) = 0.24 + 0.32 + 0.24 = 0.8$   
 $P(\mathcal{E}) = 1 - 0.8 = 0.2$

Therefore, the optimal decoder is

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

Decoder Table

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

**3.37. 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 see<sup>14</sup> 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 maximizes<sup>15</sup>  $P[X = x | Y = y]$ :

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

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

**Definition 3.38.** The **optimal decoder** derived in 3.37 is called the **maximum a posteriori probability (MAP) decoder**:

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

**3.39.** 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.

<sup>14</sup>We 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  $P(\mathcal{C})$ .

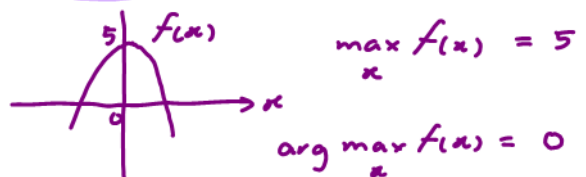
<sup>15</sup>For those who are not familiar with the “arg max” (arguments of the maximum) function,

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

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  $\max_x f(x) = 5$ .

$$\arg \max_x \frac{5 - x^2}{\sqrt{e}} = 0$$

43





## Guessing Game (for detector decoder)

Want to guess the value of  $X$

Objective: minimize the expected cost of error

$$\text{cost} = \begin{cases} 1, & \hat{X} \neq X, \\ 0, & \hat{X} = X. \end{cases}$$

$$\mathbb{E}[\text{cost}] = 1P[\hat{X} \neq X] + \cancel{0P[\hat{X} = X]} = P(\mathcal{E})$$

$$\begin{aligned} \Rightarrow \hat{x}_{\text{opt}} &= \arg \max_x p(x) \\ &= \text{the most-likely } x \text{ value} \\ &= \text{the mode of } X \\ &\quad (\text{modus}) \end{aligned}$$

For channel decoder, we can observe the value of another RV  $Y$

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

- 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  $Y = y$ .

**3.40.** We should manipulate Formula (3.38) 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 (3.38), 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  $A = [Y = y]$ .

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

Therefore, *For each  $y$ , vary the  $x$  (so look at each column)*

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

*Does not depend on  $x$  and have positive value. So, can ignore.*  
*This is the  $(x,y)$  element in the P matrix.*  
*So need to look inside the P matrix.*

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.

**3.41.** 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.
- The **corresponding  $x$  value is the value of  $\hat{x}$  for that  $y$ .**

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

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

**Example 3.42.** We have applied recipe 3.41 back when we try to find the optimal decoder in Example 3.36.

**Example 3.43.** Find the MAP decoder and its corresponding error probability for the DMC channel whose  $\mathbf{Q}$  matrix is given by

$$\mathbf{Q} = \begin{matrix} x \backslash y & 1 & 2 & 3 \\ 0 & \begin{bmatrix} 0.5 & 0.2 & 0.3 \end{bmatrix} \\ 1 & \begin{bmatrix} 0.3 & 0.4 & 0.3 \end{bmatrix} \end{matrix} \xrightarrow[\times 0.4]{\times 0.6} \mathbf{P} = \begin{matrix} x & 0 & 1 & 0 \\ y & 1 & 2 & 3 \\ \begin{bmatrix} 0.30 & 0.12 & 0.18 \\ 0.12 & 0.16 & 0.12 \end{bmatrix} \end{matrix} = \mathbf{P}$$

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

and  $\underline{\mathbf{p}} = [0.6, 0.4]$ . Note that the DMC is the same as in Example 3.26 but the input probabilities are different.

$$\hat{x}_{\text{MAP}}(y) = \begin{cases} 0, & y = 1, 3, \\ 1, & y = 2. \end{cases}$$

$$P(\mathcal{C}) = 0.3 + 0.16 + 0.18 = 0.64$$

$$P(\mathcal{E}) = 1 - 0.64 = 0.36$$

**Definition 3.44.** 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**:

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

**3.45.** 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.

$$p(x) = \begin{cases} 1/|X|, & x \in X, \\ 0, & \text{otherwise,} \end{cases} \text{ constant, positive}$$

**3.46.** 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  $\mathbf{Q}$  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  $\underline{\mathbf{p}}$  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 3.47.** Find the ML decoder and its corresponding error probability for the DMC channel in Example 3.26 whose  $\mathbf{Q}$  matrix is

$$\mathbf{Q} = \begin{array}{c|ccc} x \backslash y & 1 & 2 & 3 \\ \hline 0 & 0.5 & 0.2 & 0.3 \\ 1 & 0.3 & 0.4 & 0.3 \end{array} \xrightarrow[\text{any}]{\substack{(a) \\ (b)}} \begin{array}{c|ccc} \hat{x} & 0 & 1 & 0 \\ \hline x \backslash y & 1 & 2 & 3 \\ \hline 0 & 0.10 & 0.04 & 0.06 \\ 1 & 0.24 & 0.32 & 0.24 \end{array}$$

and  $\underline{\mathbf{p}} = [0.2, 0.8]$ .

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

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

$$P(\mathcal{E}) = 1 - 0.48 = 0.52$$

**Example 3.48.** Find the ML decoder and the corresponding error probability for a communication over BSC with  $p = 0.1$  and  $p_0 = 0.8$ .

$$\mathbf{Q} = \begin{array}{c|cc} x \backslash y & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & p & 1-p \end{array} = \begin{array}{c|cc} x \backslash y & 0 & 1 \\ \hline 0 & 0.9 & 0.1 \\ 1 & 0.1 & 0.9 \end{array} \xrightarrow[\text{any}]{\substack{(a) \\ (b)}} \begin{array}{c|cc} \hat{x} & 0 & 1 \\ \hline x \backslash y & 0 & 1 \\ \hline 0 & 0.72 & 0.08 \\ 1 & 0.02 & 0.18 \end{array}$$

$$P[x=0] = 0.8$$

$$P(\mathcal{C}) = 0.72 + 0.18 = 0.9$$

$$P(\mathcal{E}) = 1 - 0.9 = 0.1$$

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.

### 3.49. MAP vs. ML:

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

- 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  $\hat{x}_{\text{MAP}}(y)$ .

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

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

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  $\hat{x}_{\text{ML}}(y)$ .

**3.50.** In general, for BSC, it's straightforward to show that

(a) when  $p < 0.5$ , we have  $\hat{x}_{\text{ML}}(y) = y$  with corresponding  $P(\mathcal{E}) = p$ .

(b) when  $p > 0.5$ , we have  $\hat{x}_{\text{ML}}(y) = 1 - y$  with corresponding  $P(\mathcal{E}) = 1 - p$ .

(c) when  $p = 0.5$ , all four reasonable decoders have the same  $P(\mathcal{E}) = 1/2$ .

- In fact, when  $p = 0.5$ , the channel completely destroys any connection between  $X$  and  $Y$ . In particular, in this scenario,  $X \perp\!\!\!\perp Y$ . So, the value of the observed  $y$  is useless.