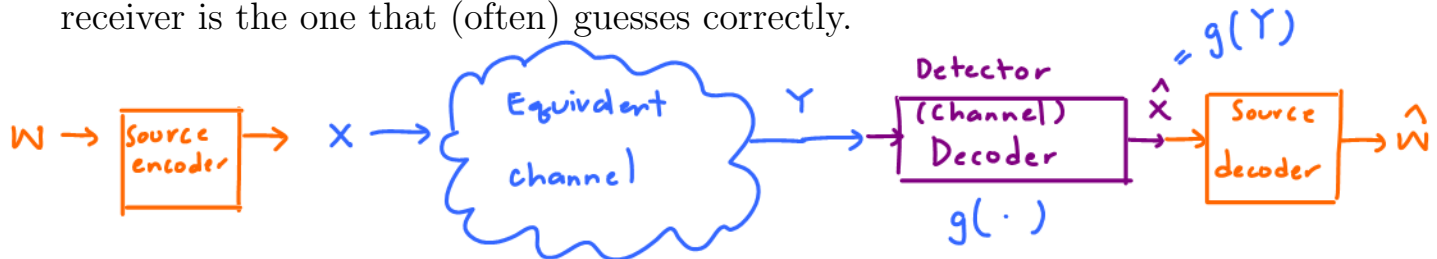## 3.2 Decoder and Symbol Error Probability

3.9. 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[11] X from the value of the received channel output Y. We denote this guessed value by $\hat{X}$. A "good" receiver is the one that (often) guesses correctly.

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

Definition 3.10. The (symbol) error probability, denoted by $P(\mathcal{E})$, can be calculated from

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

error event

3.11. A "good" detector should guess based on all the information it has. Here, the only information it receives is the value of Y. So, a detector is a function of Y, say, $g(Y)$. Therefore, $\hat{X} = g(Y)$. Sometimes, we also write $\hat{X}$ as $\hat{X}(Y)$ to emphasize that the decoded value $\hat{X}$ depends on the received value Y.

pass-through
do-nothing

Definition 3.12. A "naive" decoder is a decoder that simply sets $\hat{X} = Y$.

Example 3.13. Consider the BAC channel and input probabilities specified in Example 3.8. Find $P(\mathcal{E})$ when $\hat{X} = Y$.

$\hat{X} \equiv Y$ (naive decoder)

$$P(\mathcal{E}) \equiv P\left[\hat{X} \neq X\right] \stackrel{!}{=} P[Y \neq X]$$

$$= P[Y=1, X=0] + P[Y=0, X=1]$$

$P(A \cap B) = P(A|B) P(B) \longrightarrow$

$$= P[Y=1|X=0] P[X=0] + P[Y=0|X=1] P[X=1]$$

$P(0) = P(1) = 0.5$

---

[11]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.)

$$= 0.3 \times 0.5 + 0.4 \times 0.5 = 0.35$$

$$Q = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \xrightarrow[\times 0.5]{\times 0.5} \begin{array}{c} {}^x\backslash^y \quad 0 \quad\quad 1 \\ \begin{array}{c} 0 \\ 1 \end{array} \begin{bmatrix} 0.35 & 0.15 \\ 0.2 & 0.3 \end{bmatrix} \end{array} = P$$

$$P(\mathcal{E}) = 0.15 + 0.2 = 0.35$$

29

Naive decoder $(\hat{X} \equiv Y)$

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

$$P(\mathcal{E}) \equiv P[\hat{X} \neq X] \stackrel{.}{=} P[Y \neq X] = 1 - P[Y = X] = 1 - \sum_{c} P[Y = c, X = c]$$

$$= 1 - \sum_{x} P[Y = x, X = x] = 1 - \sum_{x} P[Y = x | X = x] P[X = x]$$

$$= 1 - \sum_{x} Q(x|x) p(x)$$

Example 3.15. With the derived formula, let's revisit Example 3.8 and Example 3.13

BAC

$$P(\mathcal{E}) \stackrel{.}{=} 1 - \sum_{x} Q(x|x) p(x) = 1 - \left(\overset{0.7 \times 0.5}{Q(0|0)p(0)} + \overset{0.6 \times 0.5}{Q(1|1)p(1)} \right) = 0.35$$

Example 3.16. Find the error probability $P(\mathcal{E})$ when a naive decoder is used with a DMC channel in which $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{1,2,3\}$, $\mathbf{Q} =$
$$\begin{array}{c} 0 \\ 1 \end{array} \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$ and $\underline{p} = [0.2, 0.8]$.

Alternatively,

$$Q = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \begin{array}{c} \times 0.2 \\ \times 0.8 \end{array}$$

$$\begin{array}{ccccc} & X \backslash Y & 1 & 2 & 3 \\ 0 & & 0.1 & 0.04 & 0.06 \\ 1 & & 0.24 & 0.32 & 0.24 \end{array} = P$$
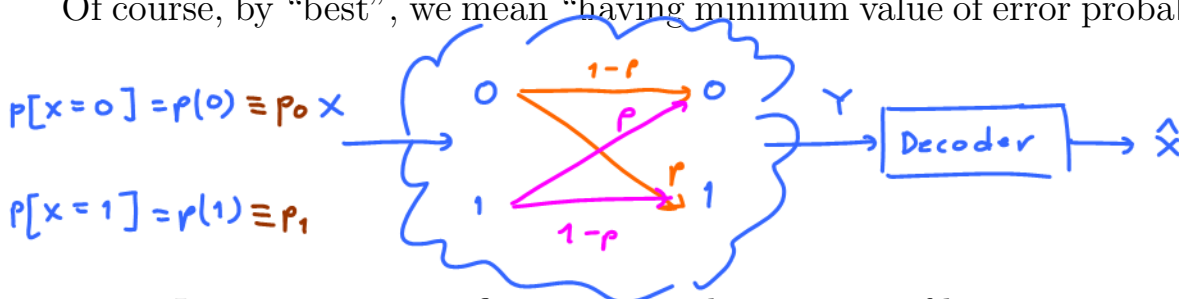
$$P(\mathcal{E}) = 1 - \sum_{x} Q(x|x) p(x)$$

$$= 1 - \left(\overset{0 \times 0.2}{Q(0|0) p(0)} + \overset{0.3 \times 0.8}{Q(1|1) p(1)} \right) = 1 - 0.24 = 0.76$$

3.3 Optimal Decoding for BSC

To introduce the idea of optimal decoding, let's revisit the binary symmetric channel in Example 3.1. Here, we will attempt to find the "best" decoder. Of course, by "best", we mean "having minimum value of error probability".

$$P[x=0] = p(0) \equiv p_0 \times$$

$$P[x=1] = p(1) \equiv p_1$$

3.17. It is interesting to first consider the question of how many reasonable decoders we can use.

Decoder uses the Y value to guess the X value.
Need to answer two questions:

① What should be the guess of X value when $Y=0$ is observed?

$\hat{X}=0$
or
$\hat{X}=1$

②   "   "   " $Y=1$ " $? \to \hat{X}=0$
or
$\hat{X}=1$

In conclusion, we need to fill out this table:

$$\begin{array}{c|c} Y & \hat{X} \\ \hline 0 & \\ 1 & \end{array}$$

So, only four reasonable detectors.

Four extreme cases:

① $\begin{array}{c|c} Y & \hat{X} \\ \hline 0 & 0 \\ 1 & 1 \end{array}$ Naive $\hat{X} \equiv Y$

② $\begin{array}{c|c} Y & \hat{X} \\ \hline 0 & 1 \\ 1 & 0 \end{array}$ $\hat{X} \equiv \overline{Y} = 1 - Y$

③ $\begin{array}{c|c} Y & \hat{X} \\ \hline 0 & 0 \\ 1 & 0 \end{array}$ $\hat{X} \equiv 0$

④ $\begin{array}{c|c} Y & \hat{X} \\ \hline 0 & 1 \\ 1 & 1 \end{array}$ $\hat{X} \equiv 1$

Example 3.18. Suppose $p = 0$. Which detector should we use?

$\hat{X}_{opt} \equiv Y$

↳ The channel does not corrupt its input: $Y \equiv X$

So, by setting $\hat{X} \equiv Y$, we always have $\hat{X} \equiv X$.

Example 3.19. Suppose $p = 1$. Which detector should we use?

$\hat{X}_{opt} \equiv \overline{Y}$

↳ The channel always flips its input value: $Y \equiv \overline{X}$

So, by setting $\hat{X} \equiv \overline{Y}$, we always have $\hat{X} = \overline{(\overline{X})} = X$

Example 3.20. Suppose $p_0 = 0$. Which detector should we use?

$\hat{X}_{opt} \equiv 1$ $P[X = 0] = 0$
$P[X = 1] = 1$

→ The channel input X is always $X \equiv 1$.

So, by setting $\hat{X} \equiv 1$, we always have $\hat{X} = X$.

Example 3.21. Suppose $p_0 = 1$. Which detector should we use?

$\hat{X}_{opt} \equiv 0$ $P[X = 0] = 1$

Example 3.22. Suppose $p = 0.1$ and $p_0 = 0.8$. Which detector should we use?

Detector	$P(\mathcal{E}) = P[\hat{X} \neq X]$
$\hat{X} = Y$	$P[Y \neq X] = 0.1$
$\hat{X} = \overline{Y}$	$P[\overline{Y} \neq X] = 1 - 0.1 = 0.9$
$\hat{X} \equiv 1$	$P[X = 0] = 0.8$
$\hat{X} \equiv 0$	$P[X = 1] = 0.2$

$p(0) = 0.8$
$p(1) = 0.2$

[diagram: x to Y channel with values 0.9, 0.1, 0.1, 0.9]
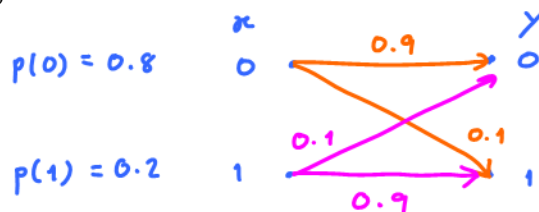
This $P(\mathcal{E})$ is the smallest among the four numbers.
So, $\hat{X} = Y$ is the optimal detector for this example.

3.23. To formally compare the performance of the four detectors, we now derive the formula for the error probability of the four detectors. First, we

31

apply the total probability theorem by using the events $[X = x]$ to partition the sample space.

$$P(\mathcal{C}) = \sum_x P(\mathcal{C} \,|[X = x]) \, P[X = x]$$

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

$$P(\mathcal{C}) = \sum_x P\left[\hat{X} = X \,|X = x\right] P[X = x] = \sum_x P\left[\hat{X} = x \,|X = x\right] P[X = x].$$

For binary channel, there are only two possible value of x: 0 or 1. So,

$$P(\mathcal{C}) = \sum_{x \in \{0,1\}} P\left[\hat{X} = x \,|X = x\right] P[X = x]$$

$$= P\left[\hat{X} = 0 \,|X = 0\right] P[X = 0] + P\left[\hat{X} = 1 \,|X = 1\right] P[X = 1]$$

$$= P\left[\hat{X} = 0 \,|X = 0\right] p_0 + P\left[\hat{X} = 1 \,|X = 1\right] (1 - p_0)$$

$\hat{X}$	$P\left[\hat{X} = 0 \,|X = 0\right]$	$P\left[\hat{X} = 1 \,|X = 1\right]$	$P(\mathcal{C})$	$P(\mathcal{E})$
Y	$1 - p$	$1 - p$	$1 - p$	p
$1 - Y$	p	p	p	$1 - p$
1	0	1	$1 - p_0$	p_0
0	1	0	p_0	$1 - p_0$

3.4 Optimal Decoding for DMC

Example 3.24. Let's return to Example 3.16 and find the error probability $P(\mathcal{E})$ when a specific decoder is used. In that example, we have $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{1, 2, 3\}$, $\mathbf{Q} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ and $\underline{p} = [0.2, 0.8]$. The decoder table below specifies the decoder under consideration:

DIY Decoder

Y	$\hat{X}$
1	0
2	1
3	0

$P(\mathcal{C}) = P[\hat{X} = x] = \sum_x P[\hat{X} = x | X = x] P[X = x]$

$= P[\hat{X} = 0 | X = 0] P[X = 0] + P[\hat{X} = 1 | X = 1] P[X = 1]$

$\qquad\qquad\qquad\quad 0.2 \qquad\qquad\qquad\qquad\qquad\quad 0.8$

$= \left(Q(1|0) + Q(3|0) \right) \times 0.2 + Q(2|1) \times 0.8$

$= 0.2 (0.5 + 0.3) + 0.4 \times 0.8 = 0.48$

$\qquad\qquad 0.1 + 0.06 + 0.32$

32

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

$G = \begin{bmatrix} 0.5 & 0.2 & 0.5 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$

$\begin{bmatrix} 0.1 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix}$

Using MATLAB, we can find the error probability for all possible reasonable detectors in this example.

3.25. For general DMC, it would be tedious to list all possible detectors. In fact, there are $|\mathcal{X}|^{|\mathcal{Y}|}$ reasonable detectors.

It is even more time-consuming to try to calculate the error probability for all of them. Therefore, in this section, we will derive the formula of the "optimal" detector.

3.26. 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:

$$P\left(\mathcal{C}\right) = \sum_{y} P\left(\mathcal{C} \,|[Y = y]\right) P\left[Y = y\right].$$

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

$$P\left(\mathcal{C}\right) = \sum_{y} P\left[\hat{X} = X \,|Y = y\right] P\left[Y = y\right].$$

Now, recall that our detector $\hat{X}$ is a function[12] of Y; that is $\hat{X} = g(Y)$ for some function g. So,

$$P\left(\mathcal{C}\right) = \sum_{y} P\left[g\left(Y\right) = X \,|Y = y\right] P\left[Y = y\right]$$

$$= \sum_{y} P\left[X = g\left(y\right) |Y = y\right] P\left[Y = y\right]$$

In this form, we see[13] that for each $Y = y$, we should maximize $P\left[X = g\left(y\right) |Y = y\right]$. Therefore, for each y, the decoder $g(y)$ should output the value of x which maximizes $P\left[X = x|Y = y\right]$:

$$g_{\text{optimal}}\left(y\right) = \arg\max_{x} P\left[X = x \,|Y = y\right].$$

---

[12]This change of notation allows one to see the dependence of $\hat{X}$ on Y.

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

In other words,

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

and

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

Definition 3.27. The *optimal* detector is the detector that maximizes the a posteriori probability $\overline{P[X = x | Y = y]}$. This detector is called the **maximum a posteriori probability (MAP) detector**:

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

- After the fact, it is quite intuitive that this should be the best detector. Recall that the decoder don't have a direct access to the X value.

    - 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 $P[X = x]$. $\hat{x}_{opt} = arg\ max_x\ p(x)$

    - Knowing $Y = y$, the decoder can update its probability about x from $P[X = x]$ to $P[X = x | Y = y]$. Therefore, the detector should guess the value of the most likely x value conditioned on the fact that $Y = y$. $\hat{x}_{opt}(y) = arg\ max_x\ P[X=x|Y=y]$

3.28. We can "simplify" the formula for the MAP detector even further.

Note that we only know $p(x),\ Q(y|x)$.
However, in the definition of the MAP detector, we need $P[X=x|Y=y]$

Fist, recall "Form 1" of the Bayes' theorem:

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

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

$$\underbrace{P[X=x|Y=y]}_{B\ A} = \underbrace{P[Y=y|X=x]\frac{P[X=x]}{P[Y=y]}}$$

$$\hat{x}_{MAP}(y) = arg\ max_x \quad = arg\ max_x \quad = arg\ max_x\ \underbrace{P[Y=y|X=x]}_{Q(y|x)}\underbrace{P[X=x]}_{p(x)}$$

34

Therefore,

$$\hat{x}_{opt}(y) = \boxed{\hat{x}_{\text{MAP}}(y) = \arg\max_x \overbrace{Q(y\,|x)\,p(x)}^{P[X=x,Y=y]}}. \tag{6}$$

3.29. A recipe to find the MAP detector and its corresponding error probability:

(a) Find the **P** matrix by scaling elements in each row of the **Q** matrix by their corresponding prior probability $p(x)$.

(b) 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 only one.

   (i) The corresponding x value is the value of $\hat{x}$ for that y.

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

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

Example 3.30. Find the MAP detector and its corresponding error probability for the DMC channel in Example 3.24 (and Example 3.16) when the prior probability vector is $\underline{p} = [0.2, 0.8]$.

$$Q = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \xrightarrow[\times 0.8]{\times 0.2} \quad \begin{array}{c} {}^{x}{\searrow}^{y} \ \ 1 \ \ \ \ 2 \ \ \ \ \ 3 \\ 0 \\ 1 \end{array} \begin{bmatrix} 0.1 & 0.04 & 0.06 \\ 0.24 & 0.32 & 0.24 \end{bmatrix} = P$$

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

$\Rightarrow \hat{x}_{\text{MAP}}(y) \equiv 1$

$P(\mathcal{C}) = 0.24 + 0.32 + 0.24 = 0.8$

$P(\mathcal{E}) = 1 - 0.8 = 0.2$

Example 3.31. Repeat Example 3.30 but with $\underline{p} = [0.6, 0.4]$.

$$Q = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \xrightarrow[\times 0.4]{\times 0.6} \quad \begin{array}{c} {}^{x}{\searrow}^{y} \ \ 1 \ \ \ \ 2 \ \ \ \ \ 3 \\ 0 \\ 1 \end{array} \begin{bmatrix} 0.3 & 0.12 & 0.18 \\ 0.12 & 0.16 & 0.12 \end{bmatrix} = P$$

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

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

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

3.32. MAP detector for BSC: For BSC,

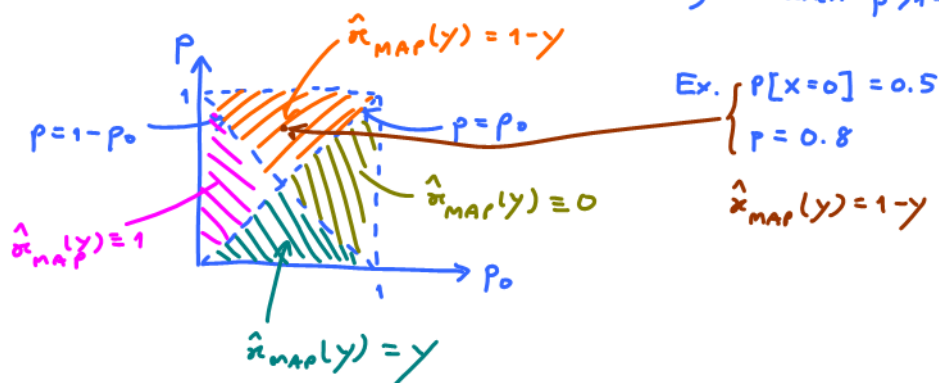$$\hat{x}_{\text{MAP}}\left(y\right) = \arg \max_{x \in \{0,1\}} Q\left(y \,|\, x\right) p\left(x\right).$$

Therefore,

$$\hat{x}_{\text{MAP}}\left(0\right) = \begin{cases} 1, & \text{when } Q\left(0\,|1\right) p\left(1\right) \overset{>}{=} Q\left(0\,|0\right) p\left(0\right), \\ 0, & \text{when } Q\left(0\,|1\right) p\left(1\right) < Q\left(0\,|0\right) p\left(0\right), \end{cases}$$

and

$$\hat{x}_{\text{MAP}}\left(1\right) = \begin{cases} 1, & \text{when } Q\left(1\,|1\right) p\left(1\right) > Q\left(1\,|0\right) p\left(0\right), \\ 0, & \text{when } Q\left(1\,|1\right) p\left(1\right) < Q\left(1\,|0\right) p\left(0\right). \end{cases}$$

Handwritten annotations:
Over first equation: $p\,(1-p_0)$, $(1-p)\,p_0$

$= \begin{cases} 1, & \text{when } p > p_0 \\ \text{any}, & \text{when } p = p_0 \\ 0, & \text{when } p < p_0 \end{cases}$

$= \begin{cases} 1, & \text{when } p < 1-p_0 \\ \text{any}, & \text{when } p = 1-p \\ 0, & \text{when } p > 1-p_0 \end{cases}$

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

$\hat{x}_{MAP}(y) = 0$

$\hat{x}_{MAP}(y) = 1$

$\hat{x}_{MAP}(y) = y$

$p = 1-p_0$, $p = p_0$

Ex. $\begin{cases} P[X=0] = 0.5 \\ p = 0.8 \end{cases}$

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

Definition 3.33. In many scenarios, the MAP detector is too complicated or the prior probabilities are unknown. In such cases, we may consider using a detector that ignores the prior probability term in (6). This detector is called the **maximum likelihood (ML)** detector:

$$\hat{x}_{\text{ML}}\left(y\right) = \arg \max_{x} Q\left(y \,|\, x\right). \tag{7}$$

Observe that

- ML detector is generally sub-optimal

- ML detector is the same as the MAP detector when X is a uniform random variable.

  ○ In other words, when the prior probabilities $p(x)$ are uniform, the ML detector is optimal.

Example 3.34. Find the ML detector and the corresponding error probability for the system in Example 3.22 in which we have BSC with $p = 0.1$ and $p_0 = 0.8$.

$\hat{x}_{ML}(y) = \arg\max\limits_{x} Q(y|x)$

$\hat{x}_{ML}(0) = \arg\max\limits_{x} Q(0|x)$ $\begin{cases} Q(0|0) = 1 - 0.1 = 0.9 \quad \vee \\ Q(0|1) = 0.1 \end{cases}$

$= 0$

$\hat{x}_{ML}(1) = \arg\max\limits_{x} Q(1|x)$ $\begin{cases} Q(1|0) = 0.1 \quad \wedge \\ Q(1|1) = 1 - 0.1 = 0.9 \end{cases}$

$= 1$

Note that

$\begin{array}{c|c} y & \hat{x}_{ML}(y) \\ \hline 0 & 0 \\ 1 & 1 \end{array} \Biggr\} \Rightarrow \hat{x}_{ML}(y) = y$

- the prior probabilities p_0 (and p_1) is not used

- the ML detector and the MAP detector are the same in this example $P(\mathcal{E}) = 0.1$

   ○ ML detector can be optimal even when the prior probabilities are not uniform.

Recall that for BSC with $\hat{x}(y) = y$, the error probability $P(\mathcal{E}) = p$. So, in this example, because $\hat{x}_{ML}(y) = y$, we have $P(\mathcal{E}) = p = 0.1$.

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

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

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

(c) when $p = 0.5$, all four reasonable detectors 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.

3.36. A recipe to find the ML detector 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 only one.
- The corresponding x value is the value of $\hat{x}$ for that y.

Find P(ℰ)

(b) Find the **P** matrix by scaling elements in each row of the **Q** matrix by their corresponding prior probability $p(x)$.

(c) In the **P** matrix, select the elements corresponding to the selected positions in the **Q** matrix.

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

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

Example 3.37. Solve Example 3.34 using the recipe in 3.36.

BSC $p = 0.1$
$p_0 = 0.8$

$$Q = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \end{array} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \xrightarrow{\substack{\times 0.8 \\ \times 0.2}} \begin{bmatrix} 0.72 & 0.08 \\ 0.02 & 0.18 \end{bmatrix} = P$$

$$\begin{array}{c|c} y & \hat{x}_{ML}(y) \\ \hline 0 & 0 \\ 1 & 1 \end{array} \qquad \hat{x}_{ML}(y) = y$$

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

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

Example 3.38. Find the ML detector and its corresponding error probability for the DMC channel in Example 3.24 (, Example 3.16, and Example 3.30) in which $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{1, 2, 3\}$, $Q = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ and $p = [0.2, 0.8]$.

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

$$\begin{array}{c|c} y & \hat{x}_{ML}(y) \\ \hline 1 & 0 \\ 2 & 1 \\ 3 & 0 \ (any) \end{array}$$

$P(\mathcal{C}) = 0.1 + 0.32 + 0.06 = 0.48$

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

38