Tutorial 3

1 Examples of Channel Capacity

1. **Noiseless Binary Channel**

   Suppose that we have a channel whose the binary input is reproduced exactly at the output as shown in Fig. 1.

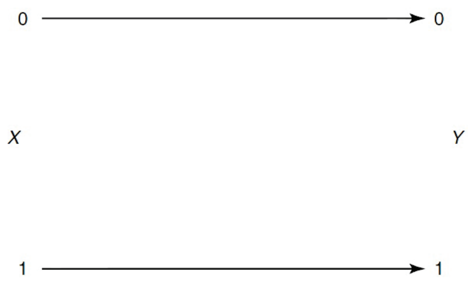


Figure 1: Noiseless Binary Channel

In this case, any transmitted bit is received without error. Hence, one error-free bit can be transmitted per use of the channel, and the capacity is 1 bit. We can also calculate the information capacity as follows.

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) - H(X|Y) = 1 \text{ bit}.$$

which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

2. **Noisy Channel with Nonoverlapping Outputs**

   This channel has two possible outputs corresponding to each of the two inputs as shown in Fig. 2.

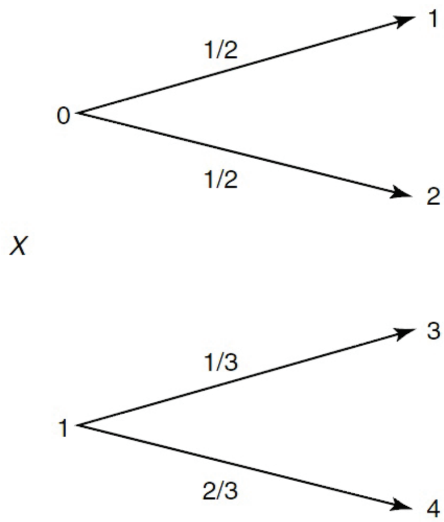


Figure 2: Noisy Channel with Nonoverlapping Outputs

The channel appears to be noisy, but really is not. Even though the output of the channel is a random consequence of the input, the input can be determined from the output, and hence every transmitted bit can be recovered without error. The capacity of this channel is also 1 bit per transmission. We can also calculate the information capacity as follows.

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) - H(X|Y) = 1 \text{ bit.}$$

3. **Noisy Typewriter**

   In this case the channel input is either received unchanged at the output with probability $\frac{1}{2}$ or is transformed into the next letter with probability $\frac{1}{2}$ as shown in Fig 3.

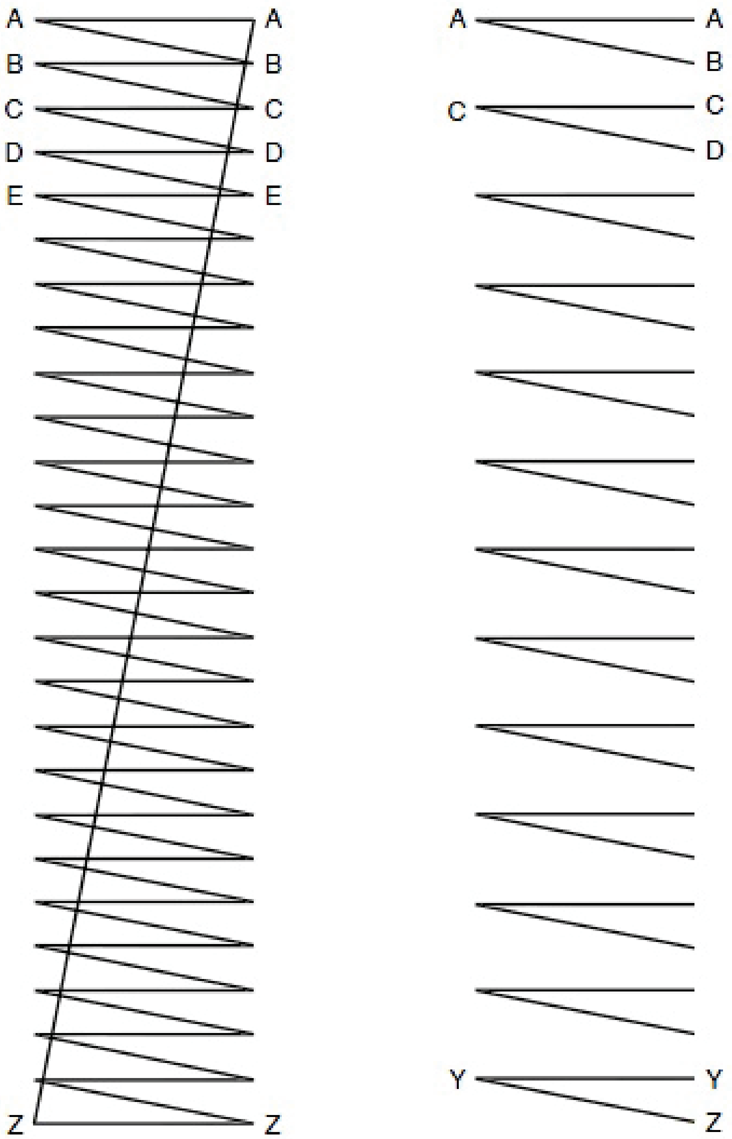


Figure 3: Noisy Typewriter

If the input has 26 symbols and we use every alternate input symbol, we can transmit one of 13 symbols without error with each transmission. Hence, the capacity of this channel is $\log 13$ bits per transmission. We can also calculate the information capacity as follows.

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(Y|X) = \log 26 - 1 = \log 13$$

which is achieved by using $p(x)$ distributed uniformly over all the inputs.

4. **Binary Symmetric Channel**

   Consider the binary symmetric channel (BSC), which is shown in Fig. 4. This is a binary channel in which the input symbols are complemented with probability p.

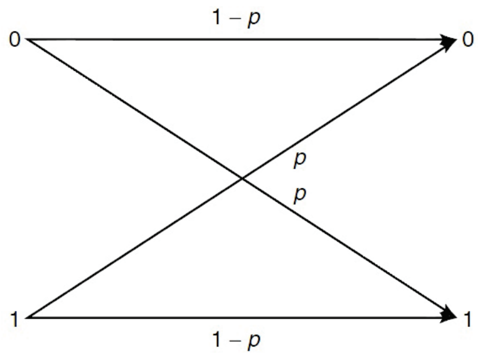


Figure 4: Binary Symmetric Channel

   We bound the mutual information by

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum p(x)H(Y|X=x) \\ &= H(Y) - \sum p(x)H(p) \\ &= H(Y) - H(p) \\ &\leq 1 - H(p) \end{aligned}$$

   where the last inequality follows because Y is a binary random variable. Equality is achieved when the input distribution is uniform. Hence, the information capacity of a binary symmetric channel with parameter p is

$$C = 1 - H(p) \text{ bits.}$$

5. **Binary Erasure Channel**

   The analog of the binary symmetric channel in which some bits are lost (rather than corrupted) is the binary erasure channel. In this channel, a fraction α of the bits are erased. The receiver knows which bits have been erased. The binary erasure channel has two inputs and three outputs as shown in Fig. 5.

   We calculate the capacity of the binary erasure channel as follows:

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - H(\alpha) \end{aligned}$$

   The first guess for the maximum of $H(Y)$ would be $\log 3$, but we cannot achieve this by any choice of input distribution $p(x)$. Letting E be the event $Y = e$, using the expansion

$$H(Y) = H(Y,E) = H(E) + H(Y|E)$$

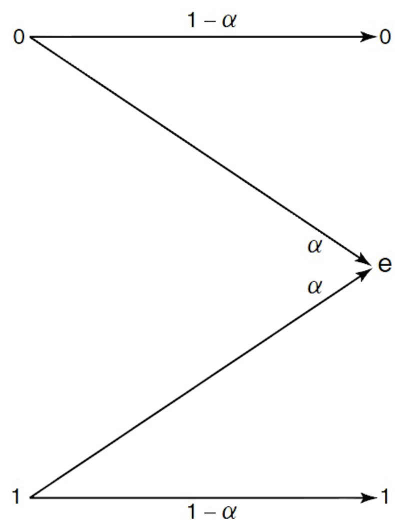


Figure 5: Binary Erasure Channel

and letting $P(X = 1) = \pi$, we have

$$\begin{aligned} H(Y) &= H((1-\pi)(1-\alpha), \alpha, \pi(1-\alpha)) \\ &= -(1-\pi)(1-\alpha)\log((1-\pi)(1-\alpha)) - \alpha\log\alpha - \pi(1-\alpha)\log(\pi(1-\alpha)) \\ &= (1-\alpha)(-(1-\pi)\log(1-\pi) - \pi\log\pi) + (-\alpha\log\alpha - (1-\alpha)\log(1-\alpha)) \\ &= (1-\alpha)H(\pi) + H(\alpha) \end{aligned}$$

Hence

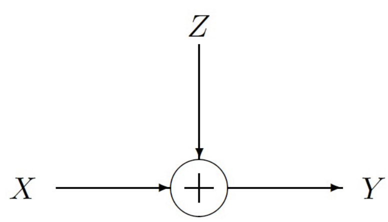
$$\begin{aligned} C &= \max_{p(x)} H(Y) - H(\alpha) \\ &= \max_{p(x)}(1-\alpha)H(\pi) + H(\alpha) - H(\alpha) \\ &= \max_{p(x)}(1-\alpha)H(\pi) \\ &= 1-\alpha \end{aligned}$$

where capacity is achieved by $\pi = \frac{1}{2}$.

2 Exercises

1. **An additive noise channel.**

   Find the channel capacity of the following discrete memoryless channel:



where $Pr\{Z = 0\} = Pr\{Z = a\} = \frac{1}{2}$. The alphabet for x is $\mathcal{X} = \{0, 1\}$. Assume the at Z is independent of X. Observe that the channel capacity depends on the value of a.

Solution:

The channel can be modeled as follows.

$$Y = X + Z \quad X \in \{0, 1\}, \quad Z \in \{0, a\}$$

We have to distinguish various cases depending on the values of a.

$a = 0$. In this case, $Y = X$, and $\max I(X;Y) = \max H(X) = 1$. Hence the capacity is 1 bit per transmission.

$a \neq 1, \pm 1$. In this case, Y has four possible values $0, 1, a$ and $1 + a$. Knowing Y, we know the X which was sent, and hence $H(X|Y) = 0$. Hence $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

$a = 1$. In this case Y has three possible output values, $0, 1$ and 2, and the channel is identical to the binary erasure channel with $a = 1/2$. Hence, the capacity of this channel is $1 - a = 1/2$ bit per transmission.

$a = -1$. This is similar to the case when $a = 1$ and the capacity is also $1/2$ bit per transmission. $\square$

2. **Channels with memory have higher capacity.**

   Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $Pr\{Z_i = 1\} = p = 1 - Pr\{Z_i = 0\}$, but that $Z_1, Z_2, \cdots, Z_n$ are not necessarily independent. Assume that Z^n is independent of the input X^n. Let $C = 1 - H(p, 1 - p)$. Show that $\max\limits_{p(x_1, x_2, \cdots, x_n)} I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n) \geq nC$.

   Solution:

   Since

$$Y_i = X_i \oplus Z_i,$$

where

$$Z_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

and Z_i are not independent,

$$\begin{aligned}
I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n) &= H(X_1, X_2, \cdots, X_n) - H(X_1, X_2, \cdots, X_n | Y_1, Y_2, \cdots, Y_n) \\
&= H(X_1, X_2, \cdots, X_n) - H(Z_1, Z_2, \cdots, Z_n | Y_1, Y_2, \cdots, Y_n) \\
&\geq H(X_1, X_2, \cdots, X_n) - H(Z_1, Z_2, \cdots, Z_n) \\
&= H(X_1, X_2, \cdots, X_n) - \sum_{i=1}^{n} H(Z_i | Z^{i-1}) \\
&\geq H(X_1, X_2, \cdots, X_n) - \sum_{i=1}^{n} H(Z_i) \\
&= H(X_1, X_2, \cdots, X_n) - nH(p) \\
&= n - nH(p)
\end{aligned}$$

if $X_1, X_2, \cdots, X_n$ are chosen i.i.d. $\sim$ Bern $\left(\frac{1}{2}\right)$. The capacity of the channel with memory over n uses of the channel is

$$\begin{aligned} nC^{(n)} &= \max_{p(x_1,x_2,\cdots,x_n)} I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n) \\ &\geq I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n)|_{p(x_1,x_2,\cdots,x_n)=\text{Bern}\left(\frac{1}{2}\right)} \\ &\geq n(1 - H(p)) \\ &= nC \end{aligned}$$

Hence channels with memory have higher capacity. □

Remark: The intuitive explanation for this result is that the correlation between the noise decreases the effective noise; one could use the information from the past samples of the noise to combat the present noise.

3. Consider the discrete memoryless channel $Y = X + Z(\text{mod}11)$, where

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3} \end{pmatrix}$$

and $X \in \{0, 1, \cdots, 10\}$. Assume that Z is independent of X.

(a) Find the capacity.

(b) What is the maximizing $p^*(x)$?

Solution:

$$Y = X + Z(\text{mod}11)$$

where

$$Z = \begin{cases} 1 & \text{with probability } \frac{1}{3} \\ 2 & \text{with probability } \frac{1}{3} \\ 3 & \text{with probability } \frac{1}{3} \end{cases}$$

In this case,

$$H(Y|X) = H(Z|X) = H(Z) = \log 3,$$

since Z is independent X, and hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - \log 3 \\ &= \log 11 - \log 3 \end{aligned}$$

which is attained when Y has a uniform distribution, which occurs (by symmetry) when X has a uniform distribution.

(a) The capacity of the channel is $\log \frac{11}{3}$ bit/transmission.

(b) The capacity is achieved by an uniform distribution on the inputs. $p(X = i) = \frac{1}{11}$ for $i = 0, 1, \cdots, 10$.

$\square$

4. **Using two channel at once.**

Consider two discrete memoryless channels $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$ with capacities C_1 and C_2 respectively. A new channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1|x_1) \times p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ is formed in which $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, are simultaneously sent, resulting in y_1, y_2. Find the capacity of this channel.

Solution:

Suppose we are given two channels, $(\mathcal{X}_1, p(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p(y_2|x_2), \mathcal{Y}_2)$, which we can use at the same time. We can define the product channel as the channel, $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$. To find the capacity of the product channel, we must find the distribution $p(x_1, x_2)$ on the input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$ that maximizes $I(X_1, X_2; Y_1, Y_2)$. Since the joint distribution

$$p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2)$$

$Y_1 \to X_1 \to X_2 \to Y_2$ forms a Markov chain and therefore

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\
&= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2, Y_1) \\
&= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
&= H(Y_1) + H(Y_2|Y_1) - H(Y_1|X_1) - H(Y_2|X_2) \\
&\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\
&= I(X_1; Y_1) + I(X_2; Y_2),
\end{aligned}$$

Equality holds if Y_1 and Y_2 are independent, namely, X_1 and X_2 are independent. Hence,

$$\begin{aligned}
C &= \max_{p(x_1, x_2)} I(X_1, X_2; Y_1, Y_2) \\
&\leq \max_{p(x_1, x_2)} I(X_1; Y_1) + I(X_2; Y_2) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\
&= C_1 + C_2
\end{aligned}$$

with equality iff $p(x_1, x_2) = p^*(x_1)p^*(x_2)$ and $p^*(x_1)$ and $p^*(x_2)$ are the distributions that maximize C_1 and C_2 respectively. $\square$

5. **The Z channel.**

The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0, 1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

Solution:

First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a funtion of $x = Pr(X = 1)$:

$$\begin{aligned} H(Y|X) &= Pr(X=0)H(Y|X=0) + Pr(X=1)H(Y|X=1) \\ &= Pr(X=0) \cdot 0 + Pr(X=1) \cdot 1 \\ &= x \\ H(Y) &= H(Pr(Y=1)) = H(x/2) \\ I(X;Y) &= H(Y) - H(Y|X) = H(x/2) - x \end{aligned}$$

Since $I(X;Y) = 0$ when $x = 0$ and $x = 1$, the maximum mutual information is obtained for some value of x such that $0 < x < 1$.

Using elementary calculus, we determine that

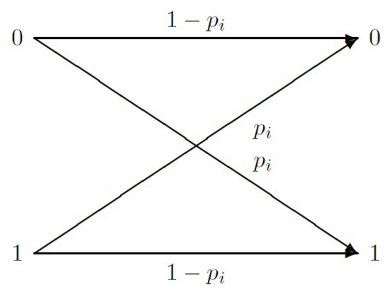
$$\frac{d}{dx} I(X;Y) = \frac{1}{2} \log \frac{1 - x/2}{x/2} - 1,$$

which is equal to zero for $x = 2/5$. (It is reasonable that $Pr(X = 1) < 1/2$ because $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel in bits is $H(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.

□

6. **Time-varying channels.**

Consider a time-varying discrete memoryless channel. Let $Y_1, Y_2, \cdots, Y_n$ be conditionally independent given $X_1, X_2, \cdots, X_n$, with conditional distribution given by $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} p_i(y_i|x_i)$. Let $\boldsymbol{X} =$



$(X_1, X_2, \cdots, X_n)$, $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$. Find $\max_{p(x)} I(\boldsymbol{X}; \boldsymbol{Y})$.

Solution:

We can use the same chain of the inequalities as in the proof of the converse of the channel coding theorem. Hence

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \cdots, Y_{i-1}, X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i), \end{aligned}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything else. Continuing the series of inequalities, we have

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\le \sum_{i=1}^{n} (1 - H(p_i)) \end{aligned}$$

with equality if $X_1, X_2, \cdots, X_n$ is chosen i.i.d. $\sim$ Bern$(1/2)$. Hence

$$\max_{p(x)} I(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n) = \sum_{i=1}^{n} (1 - H(p_i)).$$

□

7. **Unused symbols.**

Show that the capacity of the channel with probability transition matrix

$$P_{y|x} = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}$$

is achieved by a distribution that places zero probability on one of input symbols. What is the capacity of this channel?

Solution:

Let the probabilities of the three input symbols be p_1, p_2 and p_3. Then the probabilities of the three output symbols can be easily calculated to be $(\frac{2}{3}p_1 + \frac{1}{3}p_2, \frac{1}{3}, \frac{1}{3}p_2 + \frac{2}{3}p_3)$, and therefore

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(\frac{2}{3}p_1 + \frac{1}{3}p_2, \frac{1}{3}, \frac{1}{3}p_2 + \frac{2}{3}p_3) - (p_1 + p_3)H(\frac{2}{3}, \frac{1}{3}) - p_2 \log 3 \\ &= H(\frac{1}{3} + \frac{1}{3}(p_1 - p_3), \frac{1}{3}, \frac{1}{3}(p_1 - p_3)) - (p_1 + p_3)H(\frac{2}{3}, \frac{1}{3}) - (1 - p_1 - p_3) \log 3 \end{aligned}$$

where we have substituted $p_2 = 1 - p_1 - p_3$. Now if we fix $p_1 + p_3$, then the second and third term are fixed, and the first term is maximized if $p_1 - p_3 = 0$, i.e., if $p_1 = p_3$.

Now setting $p_1 = p_3$, we have

$$\begin{aligned} I(X;Y) &= H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) - (p_1 + p_3)H(\frac{1}{3}, \frac{2}{3}) - (1 - p_1 - p_3) \log 3 \\ &= \log 3 - (p_1 + p_3)H(\frac{1}{3}, \frac{2}{3}) - (1 - p_1 - p_3) \log 3 \\ &= (p_1 + p_3)(\log 3 - H(\frac{1}{3}, \frac{2}{3})) \end{aligned}$$

which is maximized if $p_1 + p_3$ is as large as possible (since $\log 3 > H(\frac{1}{3}, \frac{2}{3})$). Therefore the maximizing distribution corresponds to $p_1 + p_3 = 1$, $p_1 = p_3$, and therefore $(p_1, p_2, p_3) = (\frac{1}{2}, 0, \frac{1}{2})$. We have the capacity of this channel as

$$C = \log 3 - H(\frac{1}{3}, \frac{2}{3}) = \log 3 - (\log 3 - \frac{2}{3}) = \frac{2}{3} \text{ bits.}$$

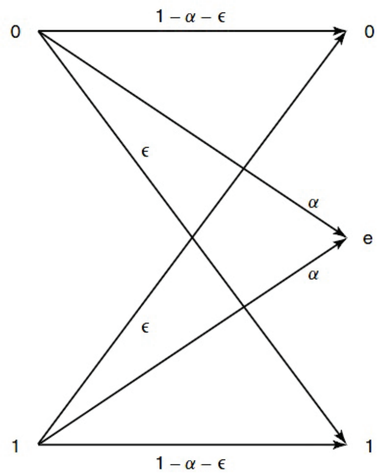
$\square$

Remark: The intuitive reason why $p_2 = 0$ for the maximizing distribution is that conditional on the input being 2, the output is uniformly distributed. The same uniform output distribution can be achieved without using the symbol 2 (by setting $p_1 = p_3$), and therefore the use of symbol 2 does not add any information (it does not change the entropy of the output and the conditional entropy $H(Y|X = 2)$ is the maximum possible, i.e., $\log 3$, so any positive probability for symbol 2 will only reduce the mutual information.

Note that not using a symbol is optimal only if the uniform output distribution can be achieved without use of that symbol. For example, in the Z channel example above, both symbols are used, even though one of them gives a conditionally uniform distribution on the output.

8. **Erasures and errors in a binary channel.**

Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ϵ and the probability of erasure be α, so the the channel is as illustrated below: (a) Find the capacity of



this channel.

(b) Specialize to the case of the binary symmetric channel ($\alpha = 0$).

(c) Specialize to the case of the binary erasure channel ($\epsilon = 0$).

Solution:

(a) As with the examples in the text, we set the input distribution for the two inputs to be π and $1 - \pi$.

Then

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} H(Y) - H(Y|X) \\
&= \max_{p(x)} H(Y) - H(1-\epsilon-\alpha, \alpha, \epsilon)
\end{aligned}
$$

As in the case of the erasure channel, the maximum value for $H(Y)$ cannot be $\log 3$, since the probability of the erasure symbol is α independent of the input distribution. Thus,

$$
\begin{aligned}
H(Y) &= H(\pi(1-\epsilon-\alpha) + (1-\pi)\epsilon, \alpha, (1-\pi)(1-\epsilon-\alpha) + \pi\epsilon) \\
&= H(\alpha) + (1-\alpha) H\left(\frac{\pi+\epsilon-\pi\alpha-2\pi\epsilon}{1-\alpha}, \frac{1-\pi-\epsilon+2\epsilon\pi-\alpha+\alpha\pi}{1-\alpha}\right) \\
&\le H(\alpha) + (1-\alpha)
\end{aligned}
$$

with equality iff $\frac{\pi+\epsilon-\pi\alpha-2\pi\epsilon}{1-\alpha} = \frac{1}{2}$, which can be achieved by setting $\pi = \frac{1}{2}$.

Therefore the capacity of this channel is

$$
\begin{aligned}
C &= H(\alpha) + 1 - \alpha - H(1-\alpha-\epsilon, \alpha, \epsilon) \\
&= H(\alpha) + 1 - \alpha - (1-\alpha) H(\frac{1-\alpha-\epsilon}{1-\alpha}, \frac{\epsilon}{1-\alpha}) \\
&= (1-\alpha)\left(1 - H(\frac{1-\alpha-\epsilon}{1-\alpha}, \frac{\epsilon}{1-\alpha})\right)
\end{aligned}
$$

(b) Setting $\alpha = 0$, we get

$$C = 1 - H(\epsilon),$$

which is the capacity of the binary symmetric channel.

(c) Setting $\epsilon = 0$, we get

$$C = 1 - \alpha,$$

which is the capacity of the binary erasure channel. $\square$