MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science
Department of Mechanical Engineering

**6.050J/2.110J** **Information and Entropy** **Spring 2005**

**Issued:** March 14, 2005

# Problem Set 7 Solutions

**Due: March 18, 2005**

## Solution to Problem 1: Confusius Gate..

### Solution to Problem 1, part a.

If each of the inputs is equally likely, then we just need to sum the probabilities of the inputs that converge at each output to find the output probabilities. Thus:

$$p(B_0) = 0.5$$
$$p(B_1) = 0.5$$

The formula for the input information is given in the lecture notes as: $I = -\sum_i p(A_i) \log P(A_i)$. the input information is then equal to

$$\begin{aligned}
I =& p(A_{00}) \log_2\left(\frac{1}{p(A_{00})}\right) + p(A_{01}) \log_2\left(\frac{1}{p(A_{01})}\right) + \\
&+ p(A_{10}) \log_2\left(\frac{1}{p(A_{10})}\right) + p(A_{11}) \log_2\left(\frac{1}{p(A_{11})}\right) \\
=& 4 \times (0.25) \log_2\left(\frac{1}{0.25}\right) \\
=& \log_2(4) \\
=& 2 \text{ bits}
\end{aligned} \tag{7–1}$$

The output information $J$ is given by the same equation

$$\begin{aligned}
I =& p(B_0) \log_2\left(\frac{1}{p(B_0)}\right) + p(B_1) \log_2\left(\frac{1}{p(B_1)}\right) \\
=& (0.5) \log_2\left(\frac{1}{0.5)}\right) + (0.5) \log_2\left(\frac{1}{0.5}\right) \\
=& 1 \text{ bit}
\end{aligned} \tag{7–2}$$

The noise $N$ is calculated with $N = -\sum_i P(A_i) \sum_j c_{ji} \log c_{ji}$ . Since $c_{ij}$ is either 1 or 0, that means that each $c_{ij} \log_2\left(\frac{1}{c_{ij}}\right)$ term is zero by virtue of the logarithm (in the case of $c_{ij} = 1$) or by $c_{ij} = 0$. Thus the noise, $N$ is zero.
The loss $L = N - J + I$, with $N = 0$:

$$\begin{aligned}
L =& I - J \\
=& 2 - 1 \\
=& 1 \text{ bit}
\end{aligned} \tag{7–3}$$

The mutual information $M$ is equal to $J$ when there is no noise.

## Solution to Problem 1, part b.

Figure 7–1 is a box diagram for the gate with defects.

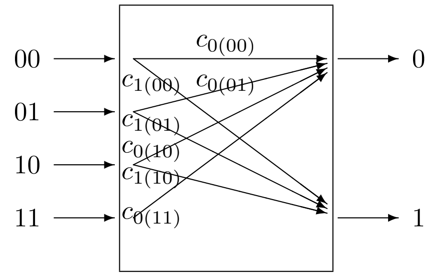


Figure 7–1: Confusius gate with defects

and the following is the transition matrix:

$$\begin{bmatrix} c_{0(00)} & c_{0(01)} & c_{0(10)} & c_{0(11)} \\ c_{1(00)} & c_{1(01)} & c_{1(10)} & c_{1(11)} \end{bmatrix} = \begin{bmatrix} .5 & 0.4 & .6 & 1 \\ .5 & 0.6 & .4 & 0 \end{bmatrix} \tag{7–4}$$

## Solution to Problem 1, part c.

Given that the result is 1 . . .

i. The probability that the 1 was produced by the input (0 1) is

$$\begin{aligned} \frac{c_{1(01)}}{c_{1(01)} + c_{1(10)} + c_{1(11)}} &= \frac{0.6}{0.5 + .6 + 0.4} \\ &= \frac{2}{5} \end{aligned} \tag{7–5}$$

ii. The probability that it was produced by the input (1 0) is

$$\begin{aligned} \frac{c_{1(10)}}{c_{1(01)} + c_{1(10)} + c_{1(11)}} &= \frac{.4}{0.5 + .6 + 0.4} \\ &= \frac{4}{15} \end{aligned} \tag{7–6}$$

iii. The probability that it was produced by the input (1 1) is

$$\frac{c_{1(11)}}{c_{1(01)} + c_{1(10)} + c_{1(11)}} = 0 \tag{7–7}$$

iv. The probability that it was produced by the input (1 0) is

$$\begin{aligned} \frac{c_{1(00)}}{c_{1(01)} + c_{1(10)} + c_{1(11)}} &= \frac{.5}{0.5 + .6 + 0.4} \\ &= \frac{1}{3} \end{aligned} \tag{7–8}$$

## Solution to Problem 1, part d.

The input information $I$ is the same as before, i.e., 2 bits. To calculate the output information first we calculate $B_0$ and $B_1$ by noting that each input is equally likely

$$\begin{aligned}
B_0 =& c_{0(00)} P(A_{00} P + c_{0(01)} P(A_{01}) + c_{0(10)} P(A_{10}) + c_{0(11)} P(A_{11}) \\
=& \frac{1}{4} \times (.5 + .4 + .6 + .1) \\
=& \frac{2.5}{4} \\
=& \frac{5}{8} && \text{(7–9)} \\
B_1 =& c_{1(00)} P(A_{00}) + c_{1(01)} P(A_{01}) + c_{1(10)} P(A_{10}) + c_{1(11)} P(A_{11}) \\
=& \frac{1}{4} \times (.5 + .6 + .4 + 0) \\
=& \frac{1.5}{4} \\
=& \frac{3}{8} && \text{(7–10)} \\
& && \text{(7–11)}
\end{aligned}$$

Then the output information $J$ is computed as before:

$$\begin{aligned}
J =& \sum_j p(B_j) \log_2 \left( \frac{1}{p(B_j)} \right) \\
=& \frac{5}{8} \log_2 \left( \frac{8}{5} \right) + \frac{3}{8} \log_2 \left( \frac{8}{3} \right) \\
=& .95 && \text{(7–12)}
\end{aligned}$$

## Solution to Problem 1, part e.

The process is both noisy and lossy, since there both fan-outs from the inputs and fan-ins to the outputs. The noise $N$ is:

$$\begin{aligned}
N =& \sum_i p(A_i) \sum_j c_{ji} \log_2 \left( \frac{1}{c_{ji}} \right) \\
=& p(A_{00}) \sum_j c_{j(00)} \log_2 \left( \frac{1}{c_{j(00)}} \right) + p(A_{01}) \sum_j c_{j(01)} \log_2 \left( \frac{1}{c_{j(01)}} \right) + \\
& + p(A_{10}) \sum_j c_{j(10)} \log_2 \left( \frac{1}{c_{j(10)}} \right) + p(A_{11}) \sum_j c_{j(11)} \log_2 \left( \frac{1}{c_{j(11)}} \right) \\
=& p(A_{00}) \left( c_{0(00)} \log_2 \left( \frac{1}{c_{0(00)}} \right) + c_{1(00)} \log_2 \left( \frac{1}{c_{1(00)}} \right) \right) + p(A_{01}) \left( c_{0(01)} \log_2 \left( \frac{1}{c_{0(01)}} \right) + c_{1(01)} \log_2 \left( \frac{1}{c_{1(01)}} \right) \right) + \\
& + p(A_{10}) \left( c_{1(10)} \log_2 \left( \frac{1}{c_{1(10)}} \right) \right) + p(A_{11}) \left( c_{0(11)} \log_2 \left( \frac{1}{c_{0(11)}} \right) + c_{1(11)} \log_2 \left( \frac{1}{c_{1(11)}} \right) \right) \\
=& \frac{1}{4} \times 1 + \frac{1}{4} \times .97 + \frac{1}{4} \times .97 \\
=& 0.735 \text{ bits} && \text{(7–13)}
\end{aligned}$$

The loss $L$ is defined as:

$$
\begin{aligned}
L =& I - (J - N) \\
=& 2 - (.95 - .73) \\
=& 1.785 \text{ bits}
\end{aligned} \tag{7–14}
$$

The mutual information $M$ is defined as:

$$
\begin{aligned}
M =& I - L \\
=& J - N \\
=& 2 - 1.785 \\
=& .95 - 0.733 \\
=& 0.214 \text{ bits}
\end{aligned} \tag{7–15}
$$

### Solution to Problem 1, part f.

The difference between $J$ for the correct channel and $J$ for the channel with defects is of .05 bits. However this number is completely uninformative, because noise and loss compete, noise gets added and compensates for the information Loss so that the final number for $J$ is similar.

$M$ is much more useful because it takes into account noise and loss. It quantifies the information that is common to both, input and output.

Without any information about noise, a good general principle for designing gates is to try to minimize the loss. That is, introduce the notion of reversibility. And let error correction and detection schemes compensate for noise.

---

## Solution to Problem 2: And then came the humans, ...

In this problem the design of the communication channel is key to get all the numbers right. We are told that we will focus in a channel that has the species of the mother as input and the species of their offspring as output. However, the information we are given comes in terms of gender of each species. In order to solve the problem we will need to address the conversion of our givens into the proper partition. This illustrates one of the most difficult aspects of information theory, translating the conditions of a real situation into the communication models we have been working with.

We have to address crossing of species and fertility, in order to determine the percentage of mothers and children of each species. There are several ways of doing so for this problem. It is often a good idea to work out the details with a diagram. Figure 7–2 shows a diagram that illustrates how to setup the desired channel.

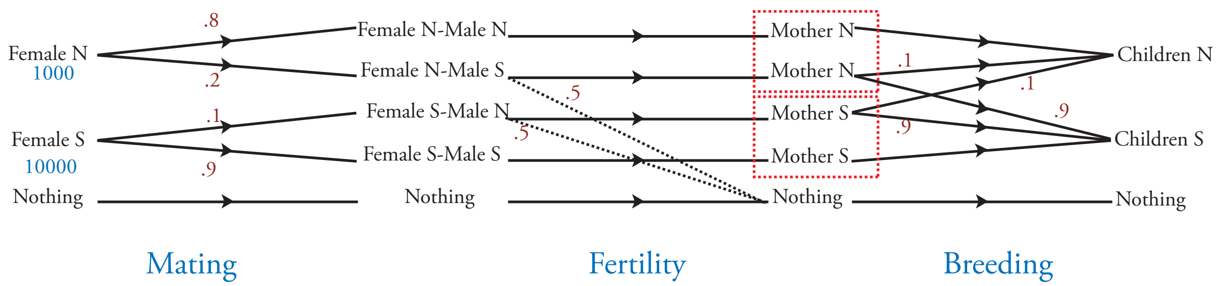


Figure 7–2: Diagram to compute the probabilities associated with the channel.

## Solution to Problem 2, part a.

Computing the probability of a randomly chosen mother being H. neanderthalis is inmediate from the diagram. We just need to count the number of mothers in the upper two branches and divide by the total number of mothers.

$$P(M_N) = \frac{.8 \times 1000 \times 1 + .5 \times .2 \times 1000}{.8 \times 1000 \times 1 + .5 \times .2 \times 1000 + .9 \times 10000 \times 1 + .5 \times .1 \times 10000} = \frac{900}{10400} = .0865 \qquad (7–16)$$

## Solution to Problem 2, part b.

This is just the input uncertainty, and we just computed everything we needed in the previous part.

$$\begin{aligned} I =& P(M_N)\log\left(\frac{1}{P(M_N)}\right) + P(M_S)\log\left(\frac{1}{P(M_S)}\right) \\ =& 0.425 \end{aligned} \qquad (7–17)$$

## Solution to Problem 2, part c.

Again, the previous diagram already contains almost all of the information we need to compute these transition probabilities. We just need to collapse each of the branches that represents mothers of the same species into one. Figure 7–3 shows the model and the transition probabilities.

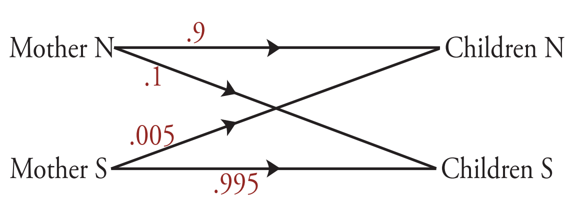


Figure 7–3: Communication model for the breeding of H. Sapiens and H. Neanderthalis.

The transition probabilities that appear in the diagram can be computed from the diagram from Figure 7–2. For example,

$$\begin{aligned} C_{NS} =& \frac{10000 \times .1 \times .5}{10000 \times .1 \times .5 + 10000 \times .9 \times .1} \times .1 \\ =& .0053 \end{aligned} \tag{7–18}$$

$$\begin{aligned} C_{SS} =& \frac{10000 \times .1 \times .5}{10000 \times .1 \times .5 + 10000 \times .9 \times 1} \times .9 + \frac{10000 \times .9 \times 1}{10000 \times .1 \times .5 + 10000 \times .9 \times 1} \\ =& .995 \end{aligned} \tag{7–19}$$

## Solution to Problem 2, part d.

We can use Bayes Rule and sum over all possible species of the mother to compute $P(C_N)$.

$$\begin{aligned} P(C_N) =& P(C_N|M_N) \times P(M_N) + P(C_N|M_S) \times P(M_S) \\ =& C_{NN} \times P(M_N) + C_{NS} \times P(M_S) \\ =& \frac{9}{10 \times 0.086 + .0053 \times .914} \\ =& .082 \end{aligned} \tag{7–20}$$

and for the children H. Sapiens

$$\begin{aligned} P(C_S) =& P(C_S|M_N) \times P(M_N) + P(C_S|M_S) \times P(M_S) \\ =& \frac{1}{10 \times 0.086 + .995 \times .914} \\ =& .92 = 1 - P(C_N) \end{aligned} \tag{7–21}$$

## Solution to Problem 2, part e.

First we need to compute the conditional probabilities of the species of the mother given the species of the children. We can use Bayes' rule to that purpose:

$$P(M_N|C_S) = \frac{P(C_S|M_N)P(M_N)}{P(C_S)} = 0.01 \tag{7–22}$$

$$P(M_S|C_S) = \frac{P(C_S|M_S)P(M_S)}{P(C_S)} = 0.99 \tag{7–23}$$

$$\tag{7–24}$$

Now we can compute the uncertainty given that the kid is H. Sapiens:

$$\begin{aligned} U_{after}(C_S) =& P(M_N|C_S) \times \log\left(\frac{1}{P(M_N|C_S)}\right) + P(M_S|C_S) \times \log\left(\frac{1}{P(M_S|C_S)}\right) \\ =& .080 \; bits \end{aligned} \tag{7–25}$$

*(Depending on the number of decimal digits you carried along your computations this number may vary)*

## Solution to Problem 2, part f.

The procedure is analogous if the kid were H. Neanderthalis,

$$P(M_N|C_N) = \frac{P(C_N|M_N)P(M_N)}{P(C_N)} = 0.942 \tag{7–26}$$

$$P(M_S|C_N) = \frac{P(C_N|M_S)P(M_S)}{P(C_N)} = 0.059 \tag{7–27}$$

$$\tag{7–28}$$

yielding the uncertainty

$$\begin{aligned} U_{after}(C_N) =& P(M_N|C_N) \times \log\left(\frac{1}{P(M_N|C_N)}\right) + P(M_S|C_N) \times \log\left(\frac{1}{P(M_S|C_N)}\right) \\ =& .32 \; bits \end{aligned} \tag{7–29}$$

## Solution to Problem 2, part g.

We just did all the calculations required for such an inference machine. Figure 7–4 shows the resulting diagram.

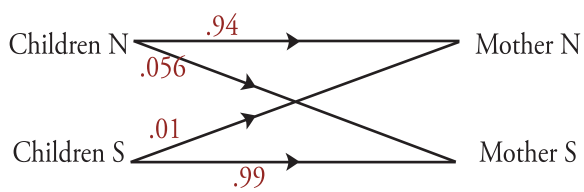


Figure 7–4: Inference Machine.

## Solution to Problem 2, part h.

Figure 7–5 illustrates the complete model of evolution. As a cascade of processes.

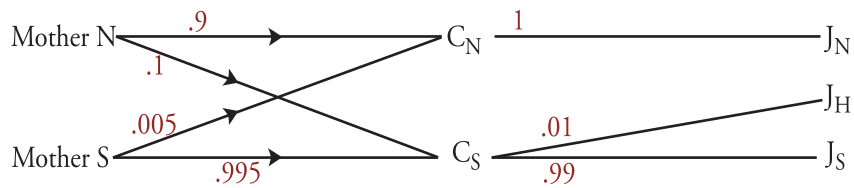


Figure 7–5: Complete model of evolution.

To characterize the mutation process we need to compute the transition probabilities. This is easily done from the information given in the text. The diagram above shows the actual result. We also need to compute input information

$$\begin{aligned} I =& P(C_S) \log\left(\frac{1}{P(C_S)}\right) + P(C_N) \log\left(\frac{1}{P(C_N)}\right) \\ =& .409\ bits \end{aligned} \tag{7–30}$$

the output distribution

$$\begin{aligned} P(J_S) &= P(C_S) \times P(J_S|C_S) \\ &= .91 \end{aligned} \tag{7–31}$$

$$\begin{aligned} P(J_H) &= P(C_S) \times P(J_H|C_S) \\ &= .01 \end{aligned} \tag{7–32}$$

$$\begin{aligned} P(J_N) &= P(C_N) \times P(J_N|C_N) \\ &= .082 \end{aligned} \tag{7–33}$$

the output information

$$J = P(J_S) \log\left(\frac{1}{P(J_S)}\right) + P(J_N) \log\left(\frac{1}{P(J_N)}\right) + P(J_H) \log\left(\frac{1}{P(J_H)}\right) = \quad .486 bits \tag{7–34}$$

We note that in this process there is no loss, therefore the mutual information will be equal to the input uncertainty:

$$M = .41$$

and the noise will be the difference $N = J - M$:

$$N = .074\ bits \tag{7–35}$$

## Solution to Problem 2, part i.

Computing the transition probabilities of the entire process, is the multiplication of the corresponding transition matrices:

$$\begin{aligned}
\mathbb{C}_{total} =& \mathbb{C}_{mutation} \times \mathbb{C}_{expected\ evolution} \\
=& \begin{bmatrix} 1 & 0 \\ 0 & .01 \\ 0 & .99 \end{bmatrix} \times \begin{bmatrix} 9/10 & .0053 \\ 1/10 & .9947 \end{bmatrix} \\
=& \begin{bmatrix} 9/10 & 0.005 \\ .001 & .00995 \\ .099 & .98505 \end{bmatrix}
\end{aligned} \tag{7–36}$$

The input and output uncertainties of the compound process, are the input uncertainty of the first process and the output uncertainty of the second process respectively.

$$I = .425 \tag{7–37}$$

$$J = .486 \tag{7–38}$$

Since the second process has no Loss, we know that the loss for the compund process is the same as the loss for the first process $L = .0971\ bits$. The mutual information is then $M = I - L = .3279\ bits$ and the noise is $N = J - M = .1581\ bits$