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 28, 2005 **Problem Set 8 Solutions** **Due: April 1, 2005**

Solution to Problem 1: Uncertain Employment

Solution to Problem 1, part a.

i. We will need to rearrange Equations 8–1 and 8–2. First we get the following from 8–2

$$P(S) = 1.2 - 2P(E) \tag{8–1}$$

and substituting $P(S)$ into 8–1 we have

$$\begin{aligned} 1 &= P(E) + P(S) + P(U) \\ &= P(E) + 1.2 - 2P(E) + P(U) \\ &= 1.2 - P(E) + P(U) \end{aligned} \tag{8–2}$$

$$P(E) = 0.2 + P(U) \tag{8–3}$$

Once $P(E)$ and $P(U)$ are determined, $P(S)$ is determined also:

$$\begin{aligned} P(S) &= 0.8 - 2P(U) \\ &= 1.2 - 2P(E) && (8–4) \\ P(U) &= P(E) - 0.2 \\ &= 0.4 - 0.5P(S) && (8–5) \\ P(E) &= 0.2 + P(U) \\ &= 0.6 - 0.5P(S) && (8–6) \end{aligned}$$

And so, since $P(U)$, $P(S)$, and $P(E)$ must be between 0 and 1, we see that $P(E)$ must be between 0.2 and 0.6. Furthermore, $P(U)$ can only be between 0 and 0.4, and $P(S)$ can only be between 0 and .8. These are reasonable answers: if the average employment is to be kept high, then we would expect that fully-employed rate would be high, but never higher than the active population. Similarly, the rate of people with no employment will be limited to be no more than the inactive population.

ii. The equation for the entropy is as follows:

$$\begin{aligned} H &= P(E)\log_2\left(\frac{1}{P(E)}\right) + P(S)\log_2\left(\frac{1}{P(S)}\right) + P(U)\log_2\left(\frac{1}{P(U)}\right) \\ &= P(E)\log_2\left(\frac{1}{P(E)}\right) + (1.2 - 2P(E))\log_2\left(\frac{1}{1.2 - 2P(E)}\right) + \\ &\quad (P(E) - 0.2)\log_2\left(\frac{1}{P(E) - 0.2}\right) \end{aligned} \tag{8–7}$$

in order to find the maximum we may either take the derivative and equal it to zero or plot in the range of valid values of $P(E)$. Taking the derivative of the formula of the entropy yields

$$\frac{\partial H}{\partial P(E)} = 2\log\left(1.2 - 2P(E)\right) - \log\left(-0.2 + P(E)\right) - \log\left(P(E)\right) \tag{8–8}$$

Equating equation 8–8 to zero and solving for $P(E)$, yields the solution: $P(E) = .438$.

iii. The maximum entropy of $H = 1.541$ bits is at $P(E) = 0.438$, which gives values of $P(S) = 0.323$ and $P(U) = 0.238$.

it is always a good idea to verify that these values satisfy our constrains. Indeed their sum is 1 and the average employment: $P(E) + .5P(S) = .6$.

Solution to Problem 1, part b.

The entropy should be less than (a-iii) because, after all, that value was calculated with a procedure known as the Principle of Maximum Entropy. The maximum value of $P(E)$ consistent with the constraints is 0.6.

Solution to Problem 1, part c.

If $P(E)$ is 0.60, then $P(S) = 0$ and $P(U) = 0.4$. The entropy at this point is $H = .971$ bits.

Solution to Problem 1, part d.

If $P(U)$ is set at its minimum level of 0, then $P(E) = 0.2$ and $P(S) = 0.8$, and the entropy of the disribution is $H = .722$.

Solution to Problem 1, part e.

The minimum entropy is zero. Any distribution that puts full weight into one of the variables will achieve zero entropy.

Solution to Problem 2: Lost in Translation

Solution to Problem 2, part a.

The probability diagram that represents the action of the editor is shown in Figure 8–1.

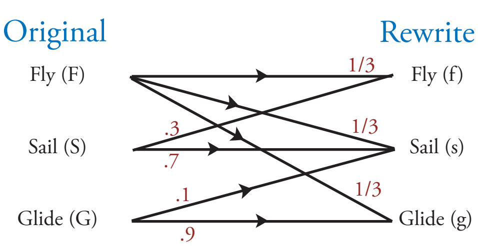

Figure 8–1: Probability diagram that represents the action of the editor

Solution to Problem 2, part b.

In general, labelling inputs as A_i and outputs as B_i, the maximum likelihood estimate is the input A_i that maximizes the likelihood $P(B_j|A_i)P(A_i)$ for a fixed B_j. One way to solve this problem is to construct a table (see Table 8–1) with outputs on the rows and inputs on the columns, that shows the likelihood for each input-output combination.

$\mathbf{P(B_i\|A_i)P(A_i)}$		Inputs A_i		
		F	S	G
Outputs B_i	f	**.3**	.015	0
	s	**.3**	.07	0
	g	**.3**	0	0

Table 8–1: Table of likelihoods. For each output B_i we search for the A_i that maximizes $P(B_i|A_i)P(A_i)$, that is the maximum by rows in the table. We have used bold font for the maximum.

As Table 8–1 shows you would assume that your spanish colleague would have always chosen fly. Which is an unrealistic choice if as an engineer he was trying to use these terms with a technical twist.

Solution to Problem 2, part c.

We can construct the same table than in part b. Table 8–2 shows that with this bias, you would assume that your british colleague would have never chosen fly. Instead every time the editor wrote fly your british colleague would really have written sail.

$\mathbf{P(B_i\|A_i)P(A_i)}$		Inputs A_i		
		F	S	G
Outputs B_i	f	.033	**.15**	0
	s	.033	**.35**	.04
	g	.033	0	**.36**

Table 8–2: Table of likelihoods. For each output B_i we search for the A_i that maximizes $P(B_i|A_i)P(A_i)$, that is the maximum by rows in the table. We have used bold font for the maximum.

Solution to Problem 2, part d.

The expression for the uncertainty is:

$$H = P(F)\log\left(\frac{1}{P(F)}\right) + P(S)\log\left(\frac{1}{P(S)}\right) + P(G)\log\left(\frac{1}{P(G)}\right). \tag{8–9}$$

if the sole constrain is that the probabilities add up to one, we have already seen in class that the maximum of entropy is achieved when all the probabilities are equal. In that case the entropy will be $H = \log 3 = 1.585$ *bits*.

Solution to Problem 2, part e.

Much like we did for the previous problem, in this part we need to express the three probabilities in terms of one of them and then find the maximum of the entropy. We can do so because we have a new constrain, namely, the average elegance.

With an average elegance of -1, we obtain the following replacement rule for G

$$P(G) \to -\frac{1}{10} + P(S) \tag{8–10}$$

Using this rule in the constrain that all probabilities add up to one we obtain a replacement rule for S

$$P(S) \rightarrow \frac{11}{20} - \frac{P(F)}{2} \tag{8–11}$$

Combining these two replacements we reahc an expression of the entropy that depends solely on $P(F)$

$$H = (10P(F) - 9)\log\left(\frac{9}{20} - \frac{P(F)}{2}\right) + (10P(F) - 11)\log\left(\frac{11}{20} - \frac{P(F)}{2}\right) - 20\ P(F)\ \log\left(P(F)\right) \tag{8–12}$$

with derivative

$$\frac{\partial H}{\partial P(F)} = \frac{\log\left(9 - 10P(F)\right) + \log\left(11 - 10P(F)\right) - 2 \times \log\left(20P(F)\right)}{\log\left(4\right)}. \tag{8–13}$$

Equating this expression to zero and using the replacements above yields the following probability distribution:

$$P(F) = 0.330829 \tag{8–14}$$

$$P(S) = 0.384586 \tag{8–15}$$

$$P(G) = 0.284586 \tag{8–16}$$

The entropy of this distribution is $H = 1.574\ bits$

Solution to Problem 2, part f.

The solution we reach by the principle of maximum entropy is still subjective, since other people could have different criteria for elegance or use notions other than elegance to constrain the distribution. What makes this solution unbiased, unlike the distributions used in parts b and c, is that all the assumptions are spelled out and the principle of maximum entropy guarantees that nothing else is unadvertently assumed.