6.050J/2.110J	**Information and Entropy**	Spring 2003

Issued: April 4, 2003   # Problem Set 8   **Due: April 11, 2003**

Note: The **quiz** will be held Thursday, April 24, 2003, 12:00 noon - 1:00 PM. The quiz will be **closed book** except that you may bring one sheet of 8 1/2 x 11 inch paper with notes on both sides. Calculators will not be necessary but you may bring one if you wish. Material through the end of Problem Set 8 may be covered on the quiz.

Problem 1: Berger's Burgers Strikes Back

Consider the example we have been following of Berger's Burgers, the restaurant. It was used in Chapter 8 of the notes to illustrate inference, in which the type of meal ordered is guessed (or at least the probabilities for each type of meal refined) as a result of knowing whether the meal arrived hot or cold. It was used in Chapter 9 to illustrate estimating probabilities compatible with constraints, by means of the Principle of Maximum Entropy (in Chapter 9 there was exactly one more unknown variable than equation so the entropy could be maximized easily). And it was also used in Chapter 10 to illustrate the Principle of Maximum Entropy with more events.

In this problem you will follow through in detail the use of this principle in a case where the maximization cannot be done easily because the entropy cannot be expressed as a function of a single variable.

You have been promoted to a middle level executive position for Berger's Burger's parent company, Carnivore Corporation, because of your great ability to estimate customer preference from a minimal amount of data. Carnivore has just acquired its competitor, Herbivore, Inc., in a hostile takeover, and is in the process of absorbing Herbivore's assets. The high-priced gourmet 200-Calorie tofu meal developed by Herbivore has been made available in all Berger's Burgers restaurants, priced at $8.

Thus your customers can select one of four meals, each with a certain price, a certain number of Calories, and a certain probability of being served cold.

Item	Entree	Cost	Calories	Probability of arriving hot	Probability of arriving cold
Value Meal 1	Burger	$1.00	1000	0.5	0.5
Value Meal 2	Chicken	$2.00	600	0.8	0.2
Value Meal 3	Fish	$3.00	400	0.9	0.1
Value Meal 4	Tofu	$8.00	200	0.6	0.4

Table 8–1: Berger's Burgers

Apparently the tofu meal is popular. You are informed by your staff nutritionist that the average Calories per meal has dropped to 380, and naturally you want to know the probability of each meal being ordered. To avoid any bias you decide to estimate the probabilities $p(B)$, $p(C)$, $p(F)$, and $p(T)$ so that the entropy, which represents your uncertainty, is as large as possible, consistent with the average Calorie count told to you.

As you start to do this problem, which is similar to one from Problem Set 7, you realize that there are now too many probabilities for the simple method used there. Specifically, there are four unknown probabilities and only two equations (the Calorie constraint and the fact that the probabilities add up to 1), so you would have to do a two-dimensional maximization. And you know that it will not be long before Carnivore expands into the wild game market with a new venison meal, and perhaps tries to attract new customers with pasta

and kosher meals. There is no end to the number of different meals you will eventually have to deal with. You need the more general procedure of Lagrange Multipliers, and you might as well start with the case of four meals.

The general procedure for the Principle of Maximum Entropy for a finite set of probabilities with two constraints is outlined here. You are asked for the particular expressions that apply in your case. For simplicity use the letters B, C, F, and T for the four probabilities $p(B)$, $p(C)$, $p(F)$, and $p(T)$.

$$\text{Probabilities} \qquad 1 \;=\; \sum_i p(A_i) \tag{8-1}$$

$$\text{Constraint} \qquad G \;=\; \sum_i g(A_i)p(A_i) \tag{8-2}$$

$$\text{Entropy} \qquad S \;=\; \sum_i p(A_i)\log_2\left(\frac{1}{p(A_i)}\right) \tag{8-3}$$

$$\text{Lagrangian} \qquad L \;=\; S - (\alpha - \log_2 e)\left(\sum_i p(A_i) - 1\right) - \beta\left(\sum_i g(A_i)p(A_i) - G\right) \tag{8-4}$$

Table 8–2: General contraint equations

$$\text{Probabilities} \qquad 1 \;=\; B + C + F + T \tag{8-5}$$

$$\text{Constraint} \qquad 380 \;=\; 1000B + 600C + 400F + 200T \tag{8-6}$$

$$\text{Entropy} \qquad S \;=\; B\log_2\left(\frac{1}{B}\right) + C\log_2\left(\frac{1}{C}\right) + F\log_2\left(\frac{1}{F}\right) + T\log_2\left(\frac{1}{T}\right) \tag{8-7}$$

$$\text{Lagrangian} \qquad L \;=\; S - (\alpha - \log_2 e)(B + C + F + T - 1)$$
$$- \beta(1000B + 600C + 400F + 200T - 380) \tag{8-8}$$

Table 8–3: Berger's Burger's contraint equations

For this to make sense, G must be larger than the smallest $g(A_i)$ and less than the greatest $g(A_i)$ (or perhaps equal to the minimum or the maximum, in which case the analysis is easy because some probabilities are zero.) The parameters introduced in the Legendre transform to get to the Lagrangian are α and β. Note that α is measured in bits, and in this case β is measured in bits per Calorie. Here e is the base of natural logarithms, 2.718282.

The Principle of Maximum Entropy is carried out by finding values of α, β, and all $p(A_i)$ such that L is maximized. These values of $p(A_i)$ also make S the largest it can be, subject to the constraints. By introducing two new variables, one for each of the constraints, we have (surprisingly) simplified the problem so that all the quantities of interest can be expressed in terms of one of the variables, and a procedure can be followed to find that one.

Since α only appears once in the expression for L, the quantity that multiplies it must be zero for the values that maximize L. Similarly, β only appears once in the expression for L, so in the general case the $p(A_i)$ that we seek must satisfy the equations

$$0 = \sum_i p(A_i) - 1 \tag{8–9}$$

$$0 = \sum_i g(A_i)p(A_i) - G \tag{8–10}$$

Answer the following questions:

a. What are the corresponding equations in your case? (Hint: these equations appear in Table 8–3 in almost the form you need.)

The result of maximizing L with respect to each of the $p(A_i)$ is

$$\log_2\left(\frac{1}{p(A_i)}\right) = \alpha + \beta g(A_i) \tag{8–11}$$

so

$$p(A_i) = 2^{-\alpha}2^{-\beta g(A_i)} \tag{8–12}$$

b. Write the corresponding equations in your case (one for each of the four meals).

Once α and β are known, the probabilities $p(A_i)$ can be found and, if desired, the entropy S can be calculated. In fact, if S is needed, it can be calculated directly, without evaluating the $p(A_i)$ – this short-cut is especially useful in cases where there are many probabilities to deal with; it is found by taking the formula for $\log_2\left(\frac{1}{p(A_i)}\right)$ above, multiplying it by $p(A_i)$, and summing over all the probabilities. The left-hand side is S and the right-hand side simplifies because α and β are independent of i. The result is

$$S = \alpha + \beta G \tag{8–13}$$

We still need to find α and β. In the general case, if Equation 8–12 is summed over the events, the result is, after a little algebra,

$$\alpha = \log_2\left(\sum_i 2^{-bg(A_i)}\right) \tag{8–14}$$

c. Derive and write the corresponding formula for your case.

We still need to find β. This is more difficult; in fact most of the computational difficulty associated with the Principle of Maximum Entropy lies in this step. In the case of just one constraint in addition to the one involving the sum of the probabilities, this step is not hard. If there are more constraints this step is more complicated.

To find β, take Equation 8–12, multiply it by $g(A_i)$ and by 2^α, and sum over i. The left hand side becomes $G2^\alpha$, because α does not depend on i. We already have an expression for α in terms of β, so the left hand side becomes $\sum_i G2^{-\beta g(A_i)}$. The right hand side becomes $\sum_i g(A_i)2^{-\beta g(A_i)}$. Thus,

$$0 = \sum_i (g(A_i) - G)2^{-\beta g(A_i)} \tag{8–15}$$

If this equation is multiplied by $2^{\beta G}$, the result is

$$0 = f(\beta) \tag{8–16}$$

where the function $f(\beta)$ is

$$f(\beta) = \sum_i (g(A_i) - G)2^{-\beta(g(A_i)-G)}. \tag{8–17}$$

    d. Derive and write the corresponding formula for $f(\beta)$ in your case.

    This is the fundamental equation that is to be solved. The function $f(\beta)$ depends on the model of the problem (the various $g(A_i)$), and on the average value G, and that is all. It does not depend on α or the probabilities $p(A_i)$. For the given value of G, the value of β that maximizes L and therefore S is the value for which $f(\beta) = 0$.

    How do we know that there is any value of β for which $f(\beta) = 0$? First, notice that since G lies between the smallest and the largest $g(A_i)$, there is at least one $g(A_i)$ for which $(g(A_i) - G)$ is positive and at least one for which it is negative. It is not difficult to show that $f(\beta)$ is a monotonic function of β, in the sense that if $\beta_2 > \beta_1$ then $f(\beta_2) < f(\beta_1)$. For large positive values of β, the dominant term in the sum is the one for the smallest value of $g(A_i)$, and hence f is negative. Similarly, for large negative values of β, f is positive. It must therefore be zero for one value of β. (For those with an interest in mathematics, this result also relies on the fact that $f(\beta)$ is a continuous function.)

    e. In your case, evaluate f for $\beta = 0$. Determine whether $f(0)$ is positive or negative (hint: it is positive).

    f. To lower f toward 0, you can increase β, for example to 0.01 bits/Calorie. Keep trying higher values until f is negative. The desired value of β lies between the last two values tried. Find β to within 0.0001 bits/Calorie. Although in principle you could use a trial-and-error procedure, you may find it much easier to use a graphing calculator or MATLAB. In fact, you might want to use MATLAB's zero-finding capabilities. The MATLAB commands are named `inline` and `fzero`. As always, to learn how to use a command, just type `help` followed by the name of the command.

    g. Now that β is known, evaluate α.

    h. Evaluate the four probabilities B, C, F, and T.

    i. Verify from those numerical values that the probabilities add up to one.

    j. Verify that these probabilities lead to the correct average number of Calories.

    k. Compute the entropy S and compare that with 2 bits, which is the number of bits needed to communicate the identity of a single meal ordered.

    l. The company treasurer wants to know what income would result from your analysis. What is the average cost per meal?

    Note that your next job might be to design the communication system from the sales counter to the kitchen. If you plan to handle many orders, then the entropy you have calculated is the amount of information your system needs to transmit per order on average. Shannon's source coding theorem states that, at least for long sequences of orders, a source coder exists that can transmit this information efficiently.

    m. Based on your calculated entropy, would it be wise to encode the stream of meal orders to transmit the information from the sales counter to the kitchen more efficiently than at the rate of 2 bits per order using a fixed-length code? Discuss the advantages and disadvantages of such a source encoding in this context. One point might be whether such a system would still work after the menu includes five items.

Turning in Your Solutions

You may turn in this problem set by e-mailing your written solutions, M-files, and diary. Do this either by attaching them to the e-mail as *text* files, or by pasting their content directly into the body of the e-mail (if you do the latter, please indicate clearly where each file begins and ends). If you have figures or diagrams you may include them as graphics files (GIF, JPG or PDF preferred) attached to your email. The deadline for submission is the same no matter which option you choose.

Your solutions are due 5:00 PM on Friday, April 11, 2003. Later that day, solutions will be posted on the course website.