CS 122A Winter 2013. HW 1 Due by 5pm Weds. January 16, 2013

Problem 1 is a warm-up question that does not require any of the material presented so far.

1. Suppose we have information about the height and weight of each person in the class, and we organize this information into a two dimensional table where there are n height intervals, and m weight intervals. In the table, the (i,j) entry indicates how many people fall into height category i and weight category j. Suppose that this table is kept secret, but the number of people falling into each height category is public, as is the number that fall into each weight category. That is, the n row sums and the m column sums of the table are published, but none of the individual cells in the table are published.

Problem 1. As above, suppose you are given a table with all cell values deleted, but the n row sums and m column sums are given to you. Call this an empty table. Assume that the sums are non-negative and that the sum of the row sums equals the sum of the column sums. Find and describe an algorithm that fills in the table with non- negative cell values so that each row and each column sums to its required amount.

1a. Explain why your algorithm never fails.

1b. Establish that an empty table can be filled in correctly (so that the entries in each row and in each colum sum to their required amounts) if and only if the sum of given row sums equals the sum of given column sums.

Answer to 1a: Let R(1),...,R(n) and C(1),...,C(m) be the positive integers labeling rows and columns respectively. A set of non-negative values for the table entries to make each row i sum to $R(i)$ and each column j sum to $C(j)$ is called a *legal solution* for the table. We will show that there is always a legal solution for the table. We do that with the algorithm shown below. As noted above, $\sum_{i=1}^{n} R(i) = \sum_{j=1}^{m} C(j)$. We denote that sum by S.

Algorithm

Set all the values in T to zero.

Set $i = 1$ and $j = 1$

Set cell (i, j) of T to $\min[R(i), C(j)]$

If the value in cell $(i, j) = R(i)$, then set $i = i + 1$, and set $C(j) = C(j) - R(j)$.

Else if the value in cell $(i, j) = C(j)$, then set $j = j + 1$, and set $R(i) = R(i) - C(j)$.

If $i = n + 1$ or $j = m + 1$ Stop.

Claim: the above algorithm always finds a legal solution.

To prove this, note that when i is incremented to $i+1$, the sum of numbers set in row i must be $R(i)$, and when j is incremented to $j + 1$, the sum of numbers set in column j must be $C(j)$. Note also that the values of i and j never decrease. So at every iteration, the requirements of one row or one column is satisfied. When the algorithm stops, either $i = n+1$ or $j = m+1$. Suppose it is $i = n + 1$ (the other case is symmetric). That implies that the values in each row p sum exactly to $R(p)$, for p from 1 to n, and the values in the table sum to exactly S. That is, when you sum the values in the table by looking along each row, the sum will be exactly S. By the same reasoning, the values in each column q sum to $C(q)$ for $q < j$, and the sum of the values in column j is less than or equal to $C(j)$. But when you sum the values in the table by looking down each column, the sum must again be S (since all the cell values in the table are summed). So, when the algorithm terminates, it must also be true that the sum of values in column q is $C(q)$ for q from 1 to m, and hence the algorithm does find a legal solution.

Next to answer part 1b, we could prove the *if* part of the statement by using the above algorithm, with the same proof of correctness. Alternatively, we could have proved the *if* part without using an algorithm, for example by the following induction on n:

Consider first a 1 by m table where the $R(1) = \sum_{j=1}^{m} C(j)$. Set each cell $1, j$ value to $C(1, j)$. Clearly, that specifies a legal solution for the 1 by m table. That is the basis for the induction proof. For the inductive hypothesis, assume there is a legal solution for tables of size up to n' by m, for some n'. For the induction step, consider a table with one more row, $n' + 1$. Set the value of cell $(n' + 1, 1)$ to $\min[R(n' + 1), C(1)]$, say $R(n' + 1)$ (the case when the Min is $C(1)$ is symmetric and omitted). Clearly, the other cells in row $n' + 1$ must then be set to zero. Now, removing row $n' + 1$ and reducing the requirement for column 1 by $R(n' + 1)$, we have an n' by m table problem with row and column sums requirements that are equal, hence which satisfy the induction hypothesis. So, there is a legal solution for that n' by m table. Put together with the values for row $R(n' + 1)$, this give a legal solution to the table of size $n' + 1$ by m, and so by the principle of induction, there is a legal solution to any table, where the row sums and the column sums have the same totals.

—

To finish the solution to problem 1b. we need to show that there is a legal solution *only if* the row sums and the column sums have the same totals. That means that the existence of a legal solution T' for table T implies that $\sum_{i=1}^{n} R(i) = \sum_{j=1}^{m} C(j)$.

We can sum all the entries in the filled-in table T', either by looking along the rows, or by looking along the columns. The total must be the same both ways. Looking along the rows, we see that the table total is $\sum_{i=1}^{n} R(i)$, and looking along the columns, the table sum is $\sum_{j=1}^{m} C(j)$. Hence, we have established that $\sum_{i=1}^{n} R(i) = \sum_{j=1}^{m} C(j)$ as required.