Ben Johnson
Problem of the Week 5

Definition: A binary matrix is even if all of its rows and columns contain an even number of ones.

Let A be an m by n binary matrix with the element of A at row i, column j designated as $a_{i,j}$. There is a nice visualization of this number of independent variables. Choose $a_{i,j}$ any way you like for $i < m$ and $j < n$. Then let

$$a_{i,n} = \sum_{j=1}^{n-1} a_{i,j}$$

$$a_{m,j} = \sum_{i=1}^{m-1} a_{i,j}$$

$$a_{m,n} = \sum_{i=1}^{m-1} a_{i,n} = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} a_{i,j} = \sum_{j=1}^{n-1}\sum_{i=1}^{m-1} a_{i,j} = \sum_{j=1}^{n-1} a_{m,j}$$

where the sums are taken over the finite field GF(2).

In this visualization, when adding any row or column, the last entry in the row or column is equal to the parity of the preceding entries, so the parity of the entire row or column must be zero. The parity of the final row minus its last entry must be equal to the parity of the final column minus its last entry as this parity is equal to the parity of the sum of the entries of the $m-1$ by $n-1$ upper-left sub matrix of A. This visualization places a lower bound of $2^{mn-m-n+1}$ on the number of even binary matrices, but it does not show that all even binary matrices can be constructed in this way. A less intuitive but more rigorous treatment follows which shows that $2^{mn-m-n+1}$ is in fact the largest possible number of binary matrices.

Let A be an m by n binary matrix with the element of A at row i, column j designated as $a_{i,j}$. Then the matrix is even if and only if the two conditions are satisfied:

$$\sum_{j=1}^{n} a_{i,j} = 0 \forall 1 \leq i \leq m \qquad (1)$$

$$\sum_{i=1}^{m} a_{i,j} = 0 \forall 1 \leq j \leq n \qquad (2)$$

where the sums are taken over the finite field GF(2). Now this is a system of linear equations in a finite field. Since all of the variables in any two of the m linear equations specified by Equation 1 must have distinct j, each equation has a distinct set of variables, so surely the row equations are linearly independent. In other words, Equation 1 specifies m linearly independent equations in GF(2), and Equation 2 specifies n linearly independent equations in GF(2).

However, if all the linear equations specified in Equation 1 are added and all the equations in Equation 2 are added, the results are the same:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} \quad = \quad \sum_{j=1}^{n}\sum_{i=1}^{m} a_{i,j}$$

Therefore, at least one of the equations is redundant. Removing the first equation specified by Equation 2, the new, equivalent system of linear equations is

$$\sum_{j=1}^{n} a_{i,j} = 0 \forall 1 \leq i \leq m \tag{3}$$

$$\sum_{i=1}^{m} a_{i,j} = 0 \forall 2 \leq j \leq n \tag{4}$$

Construct a p by q binary matrix of coeffecients C of this system of linear equations as follows. Label the first m columns $(1,1)$ to $(m,1)$, then label subsequent columns $(1,2)$ to $(1,n)$, $(2,2)$ to $(2,n)$, ..., $(m,2)$ to (m,n). Label the rows $i=1$ to $i=m$ followed by $j=2$ to $j=n$.

If a row is labelled $i=h$, then each entry in that row will contain a 1 if and only if its column is labelled (h,k) for some k. The rows labelled $i=h$ completely describe the system of equations given by Equation 3. If a row is labelled $j=k$, then each entry in that row will contain a 1 if and only if its column is labelled (h,k) for some h. These rows completely describe the system of equations given by Equation 4. So the matrix C is a matrix of coeffecients that completely describes the restrictions on the entries of A.

Now certainly, for $1 \leq h \leq m$, the corresponding row is labeled $i=h$. Then $c_{h,h}=1$ as the column is labelled $(h,1)$. Furthermore, $c_{h,k}=0$ for $1 \leq k < h$ as those columns will be labelled $(k,1)$ where $k < h$. In the case where $m < h \leq q$, the row is labelled $j=k$ for some $k > 1$. The first m entries will contain zeroes as those columns are labelled $(1,1)$ to $(m,1)$ and $k \neq 1$. The next $k-2$ entries are labelled $(1,2)$ to $(1,k-1)$, so clearly these columns are all zeroes as well. The very next column, labelled $(1,k)$, will be the first column to contain a 1. So there a total of $m+k-2$ zero entries followed by a 1 entry. But $k=2$ for the $(m+1)$th row, so the row index $h=m+k-1$. Therefore, for all rows the leading coeffecient is on the main diagonal, meaning that this specification of C is in row echelon form!

How many free variables does C have? As C is already in row echelon form, the number of free variables is the number of variables to the right of the last entry in the main diagonal, or $q-p$. With our choice of labelling of C it is clear that $p=m+n-1$ and $q=mn$, so there are $mn-m-n+1$ independent variables. These variables may assume $2^{mn-m-n+1}$ possible values, so there are $2^{mn-m-n+1}$ distinct configurations of an even m by n matrix.