

How Much Does a Matrix of Rank k Weigh?

THERESA MIGLER

California Polytechnic State University
San Luis Obispo, CA 93407
tmigler@calpoly.edu

KENT E. MORRISON

California Polytechnic State University
San Luis Obispo, CA 93407
kmorriso@calpoly.edu

MITCHELL OGLE

California Polytechnic State University
San Luis Obispo, CA 93407
mogle@calpoly.edu

Matrices with very few nonzero entries cannot have large rank. On the other hand matrices without any zero entries can have rank as low as 1. These simple observations lead us to our main question. For matrices over finite fields, what is the relationship between the rank of a matrix and the number of nonzero entries in the matrix? This question motivated a summer research project collaboration among the authors (two undergraduate students and their adviser), and although the question seems natural, we were unable to find any previously published work dealing with it.

We call the number of nonzero entries of a matrix A the *weight* of A and denote it by $\text{wt } A$. For matrices over finite fields, the weight of $A - B$ is a natural way to define the distance between A and B . In coding theory the distance between vectors defined in this way is called the *Hamming distance*, named after Richard Hamming, a pioneer in the field of error correcting codes. The rank of $A - B$, denoted $\text{rk}(A - B)$, defines a different distance between the matrices A and B . Thus $\text{wt } A$ and $\text{rk } A$ give two ways to measure the distance from A to the origin and our fundamental question is about the relationship between them.

The background needed for this paper comes from the undergraduate courses in linear algebra, abstract algebra, and probability. We use the fundamental ideas of linear algebra over finite fields. For each prime power q we let \mathbf{F}_q denote the unique field with q elements. There is no problem, however, in reading the rest of the paper with only the prime fields \mathbf{F}_p (or even \mathbf{F}_2) in mind. We use the basic concepts and results of probability up through the central limit theorem.

Having restricted our investigation to matrices over finite fields, we restate the fundamental question in this way: Over \mathbf{F}_q how many $m \times n$ matrices of rank k and weight w are there? In probabilistic terms, we are asking for the distribution of the weight for matrices of rank k . We do not have the complete answer to this question, and it seems we are far from the complete answer, so there is plenty of work left to be done. The main results we offer are the average value of the weight for matrices of fixed rank and the complete description of the weight distribution for rank 1 matrices.

Counting matrices

For matrices of fixed size, the zero matrix is the only matrix of weight 0 and the only matrix of rank 0. Also, any matrix of weight 1 has rank 1, and a matrix of rank k has

weight at least k . It is possible to count the matrices of very small rank and weight. For a matrix of rank 1 and weight 1 we choose one of the mn entries in which to place one of the $q - 1$ nonzero elements of the field. That gives us $mn(q - 1)$ matrices of rank 1 and weight 1. We leave two more results as exercises for the reader.

- The number of rank 1 and weight 2 is

$$\frac{1}{2}mn(m + n - 2)(q - 1)^2.$$

- The number of rank 2 and weight 2 is

$$\frac{1}{2}mn(m - 1)(n - 1)(q - 1)^2.$$

The number of matrices of weight w (regardless of rank) is easy to count. There are w locations to select and in each location there are $q - 1$ nonzero elements to choose. Therefore, the number of $m \times n$ matrices of weight w is

$$\binom{mn}{w}(q - 1)^w.$$

The probability that a matrix has weight w , assuming that each matrix is equally probable, is then

$$\frac{1}{q^{mn}} \binom{mn}{w} (q - 1)^w = \binom{mn}{w} (1 - 1/q)^w (1/q)^{mn-w},$$

showing us that the weight follows a binomial distribution with parameters mn and $1 - 1/q$.

Next we count the number of $m \times n$ matrices of rank k without regard to weight. Although this is a more difficult problem than counting according to weight, it is an old result with the first derivation of the formula due to Landsberg in 1893 [5]. We break the problem into two parts by first counting the matrices with a fixed column space of dimension k and then counting the subspaces of dimension k . The product of these two numbers is the number of $m \times n$ matrices of rank k .

Let V be a fixed k -dimensional subspace of the m -dimensional space \mathbb{F}_q^m . The $m \times n$ matrices whose column space is V are bijective with the linear transformations from \mathbb{F}_q^n onto V , which are bijective with the $k \times n$ matrices of rank k .

FORMULA 1. *The number of $k \times n$ matrices of rank k is*

$$\prod_{0 \leq i \leq k-1} (q^n - q^i) = (q^n - 1)(q^n - q) \cdots (q^n - q^{k-1}).$$

Proof. In order for a $k \times n$ matrix to have rank k the rows must be linearly independent. (For a slick proof that row rank equals column rank see the recent note by Wardlaw in this magazine [6].) Now the first row can be any nonzero vector with n entries, and there are $q^n - 1$ such vectors. The second row must be independent of the first row. That means it cannot be any of the q scalar multiples of that row, but any other row vector is allowed. There are $q^n - q$ vectors to choose from. The third row can be any vector not in the span of the first two rows. There are q^2 linear combinations of the first two rows, and so there are $q^n - q^2$ possible vectors for row 3. We continue in this way with row $i + 1$ not allowed to be any of the q^i linear combinations of the first i rows. ■

FORMULA 2. *The number of k -dimensional subspaces of an m -dimensional vector space over \mathbf{F}_q is*

$$\frac{\prod_{0 \leq i \leq k-1} (q^m - q^i)}{\prod_{0 \leq i \leq k-1} (q^k - q^i)}.$$

Proof. To count the number of k -dimensional subspaces of a vector space of dimension m , we count the number of bases of all such subspaces and then divide by the number of bases that each subspace has. A basis is an ordered list of k linearly independent vectors lying in \mathbf{F}_q^m . Putting them into a matrix as the rows, we get a matrix of rank k . From Formula 1 (with n replaced by m) we see that there are $\prod_{0 \leq i \leq k-1} (q^m - q^i)$ bases. The number of bases of a k -dimensional space is just the number of $k \times k$ matrices of rank k . Again, we use Formula 1 (with n replaced by k) to see that there are $\prod_{0 \leq i \leq k-1} (q^k - q^i)$ bases of a particular subspace. ■

There is a well-developed analogy in the world of combinatorics between the subsets of a finite set and the subspaces of a finite dimensional vector space over a finite field. The number of k -dimensional subspaces of an m -dimensional vector space is analogous to the number of subsets of size k in a set of size m , which is given by the binomial coefficient $\binom{m}{k}$. So, we let

$$\binom{m}{k}_q := \frac{\prod_{0 \leq i \leq k-1} (q^m - q^i)}{\prod_{0 \leq i \leq k-1} (q^k - q^i)}$$

denote the number of such subspaces, as given in Formula 2. This number is often called a *Gaussian binomial coefficient*. Although the full development of the subset-subspace analogy is not necessary for us, we recommend Kung’s introductory survey [4] and Cohn’s recent discussion of the Gaussian binomial coefficients [2].

Formulas 1 and 2 give the two factors we need for the number of $m \times n$ matrices of rank k . As one should expect the formula is symmetric in m and n .

FORMULA 3. *The number of $m \times n$ matrices of rank k is*

$$\frac{\prod_{0 \leq i \leq k-1} (q^n - q^i) \prod_{0 \leq i \leq k-1} (q^m - q^i)}{\prod_{0 \leq i \leq k-1} (q^k - q^i)}.$$

The average weight of rank k matrices

As we have mentioned, the weight of a matrix A is the Hamming distance between A and the zero matrix, and so we expect that in some way increasing weight is correlated with increasing rank. In this section, we determine the average weight of the set of matrices of a fixed rank in terms of the parameters q, m, n , and k . Indeed we find that the average weight grows with k when the other parameters are held fixed.

We consider the weight as a random variable W , which is the sum $\sum_{i,j} W_{ij}$, where W_{ij} is the weight of the i, j entry, meaning that $W_{ij} = 1$ for a matrix whose i, j entry is nonzero and $W_{ij} = 0$ when the entry is 0. Then the average or expected value of W is the sum of the expected values of the random variables W_{ij} . The expected value of W_{ij} is simply the probability that the i, j entry is nonzero. It is this probability that we will compute. An important observation is that this probability is the same for all i and j . In other words, the W_{ij} are identically distributed.

THEOREM 1. *For $m \times n$ matrices of rank k , the probability that the i, j entry is nonzero is the same for all i and j .*

Proof. For a fixed row index i and column index j there is a bijection on the space of $m \times n$ matrices defined by switching row 1 with row i and switching column 1 with column j . This bijection preserves the rank and weight, and so it defines a bijective correspondence between the subset of matrices of rank k with nonzero 1,1 entry and the subset of matrices of rank k with nonzero i, j entry. ■

With this result we know that the expected value of W is mn times the average weight of the 1,1 entry, so that we can focus our attention on the upper left entry. Our analysis depends on what is called the *reduced row echelon form*. Recall that a matrix is in reduced row echelon form if all the nonzero rows are above all the zero rows, if the leftmost nonzero entry of a nonzero row is 1, and if such an entry is the only nonzero one in its column.

The reduced row echelon forms are actually a system of representatives of the *row equivalence classes*, where two matrices are row equivalent if a sequence of elementary row operations changes one into the other. It is also the case that A and B are row equivalent if and only if they have identical row spaces.

For an $m \times n$ matrix A of rank k , the reduced row echelon form of A has k nonzero rows. Let R be the $k \times n$ matrix consisting of those rows. Since the rows of R form a basis of the row space of A , each row of A is a linear combination of the rows of R . That means there is a unique $m \times k$ matrix C such that $A = CR$. Note that the rank of C must also be k .

Using the factorization $A = CR$ we can express the set of rank k matrices as the Cartesian product of the set of $m \times k$ matrices of rank k with the set of $k \times n$ reduced row echelon matrices of rank k . This means that A can be selected randomly by independently choosing the factors C and R . Now the 1,1 entry of A is given by $a_{11} = c_{11}r_{11} + c_{12}r_{21} + \cdots + c_{1k}r_{k1}$. Since R is in reduced row echelon form, r_{11} is 0 or 1 and the rest of the entries in the first column, $r_{21}, r_{31}, \dots, r_{k1}$, are all 0. Therefore, $a_{11} = c_{11}r_{11}$, and so the probability that the 1,1 entry is nonzero is

$$\mathbf{P}(a_{11} \neq 0) = \mathbf{P}(c_{11} \neq 0)\mathbf{P}(r_{11} \neq 0).$$

Now C is $m \times k$ and has rank k . Thus, the first column of C is any nonzero vector of length m , of which there are $q^m - 1$. There are $q^{m-1} - 1$ of those vectors that have a zero in the top entry, and so there are $q^m - q^{m-1}$ that have a nonzero top entry. Then

$$\mathbf{P}(c_{11} \neq 0) = \frac{q^m - q^{m-1}}{q^m - 1}.$$

The choice of the reduced matrix R is the same as the choice of row space of A . If any of the vectors in the row space has a nonzero first entry, then the first column cannot be the zero column and then r_{11} is not 0. In order that $r_{11} = 0$ the row space of A must be entirely within the $n - 1$ dimensional subspace of vectors of the form $(0, x_2, x_3, \dots, x_n)$. The probability of that occurring is the ratio of the number of k -dimensional subspaces of a space of dimension $n - 1$ to the number of k -dimensional subspaces of a space of dimension n :

$$\mathbf{P}(r_{11} = 0) = \frac{\binom{n-1}{k}_q}{\binom{n}{k}_q}.$$

Therefore, the complementary probability gives

$$\mathbf{P}(r_{11} \neq 0) = 1 - \frac{\binom{n-1}{k}_q}{\binom{n}{k}_q}.$$

Using Formula 2 we simplify this to get

$$\mathbf{P}(r_{11} \neq 0) = \frac{q^n - q^{n-k}}{q^n - 1}.$$

Putting these results together gives us

$$\mathbf{P}(a_{11} \neq 0) = \left(\frac{q^m - q^{m-1}}{q^m - 1} \right) \left(\frac{q^n - q^{n-k}}{q^n - 1} \right).$$

As a probability this is more easily analyzed in the following form:

$$\mathbf{P}(a_{11} \neq 0) = \frac{(1 - 1/q)(1 - 1/q^k)}{(1 - 1/q^m)(1 - 1/q^n)}.$$

When m , n , and k are large, this probability is close to $1 - 1/q$, which is the probability that an entry is nonzero with no condition on the rank. One case of interest is that of invertible matrices. For $n \times n$ invertible matrices we have $k = m = n$, and so the probability that an entry is nonzero simplifies to

$$\frac{1 - 1/q}{1 - 1/q^n}.$$

We see that it is slightly more likely that an invertible matrix has nonzero entries than an arbitrary matrix.

Having determined the probability that the 1, 1 entry is nonzero and hence that the probability that the i, j entry is nonzero, we have proved the following theorem.

THEOREM 2. *The average weight of an $m \times n$ matrix of rank k over the field of order q is*

$$mn \frac{(1 - 1/q)(1 - 1/q^k)}{(1 - 1/q^m)(1 - 1/q^n)}.$$

We also see that with m and n fixed the average weight increases as k increases. It is this formula that best expresses the intuitive idea that increasing rank is correlated with increasing weight. FIGURE 1 shows a plot of the average weight vs. the rank for matrices of size 10×10 over the field \mathbf{F}_2 .

The weight of rank 1 matrices

We can analyze the weight distribution more completely for matrices of rank 1. From Theorem 2 with $k = 1$ we see that the average weight of a rank 1 matrix is

$$mn \frac{(1 - 1/q)^2}{(1 - 1/q^m)(1 - 1/q^n)}.$$

For m and n large this average is just about $mn(1 - 1/q)^2$, whereas the average weight for all $m \times n$ matrices is $mn(1 - 1/q)$, and so rank 1 matrices tend to have a lot more zero entries than the average matrix. For $q = 2$ this effect is the most pronounced. An average of one fourth of the entries are 1 in a large rank 1 matrix over \mathbf{F}_2 , while an average of half the entries are 1 for all matrices.

In the factorization $A = CR$, where $\text{rk } A = 1$, C is a nonzero column vector of length m and R is a nonzero row vector of length n whose leading nonzero entry is 1.

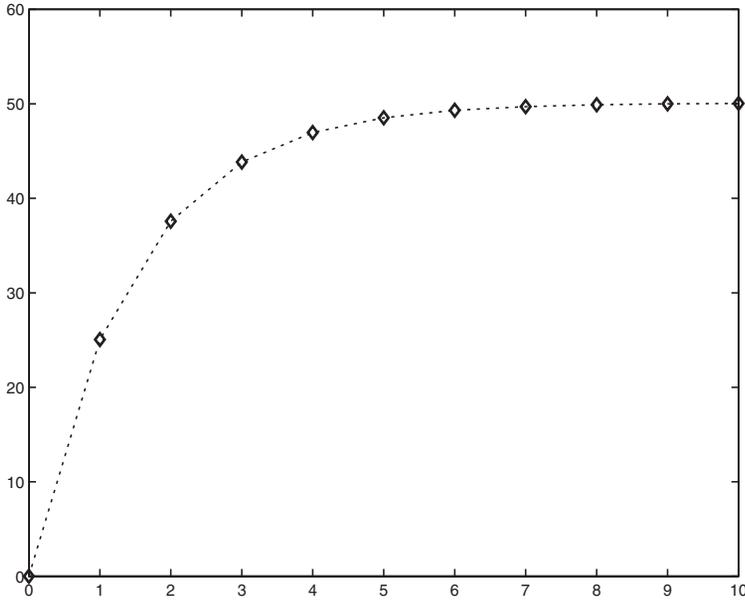


Figure 1 Average weight plotted against rank for 10×10 matrices over F_2

The entries of A are given by $a_{ij} = c_i r_j$, and so the weight of A is the product of the weights of C and R . The weight of C has probability distribution given by

$$\mathbf{P}(\text{wt } C = \mu) = \frac{\binom{m}{\mu}(q - 1)^\mu}{(q^m - 1)},$$

because there are $q^m - 1$ nonzero column vectors of length m , and there are

$$\binom{m}{\mu}(q - 1)^\mu$$

vectors of weight μ . This is the distribution of a binomial random variable (with parameters m and $1 - 1/q$) conditioned on being positive. Likewise for R the weight distribution is given by

$$\mathbf{P}(\text{wt } R = \nu) = \frac{\binom{n}{\nu}(q - 1)^\nu}{(q^n - 1)}.$$

To select a random R , choose a random nonzero vector of length n and then scale it to make the leading nonzero entry 1. The scaling does not change the weight.

Immediately we see that there is a restriction on the possible weight of a matrix of rank 1. For example, the weight of a 3×4 matrix of rank 1 cannot be 5, 7, 10, or 11 because those numbers are not products $\mu\nu$ with $1 \leq \mu \leq 3$ and $1 \leq \nu \leq 4$. All other weights between 1 and 12 are possible.

The weight of rank 1 matrices is the product of these two binomial random variables, each conditioned to be positive.

$$\begin{aligned} \mathbf{P}(\text{wt } A = \omega) &= \sum_{\mu\nu=\omega} \mathbf{P}(\text{wt } C = \mu)\mathbf{P}(\text{wt } R = \nu) \\ &= \sum_{\mu\nu=\omega} \binom{m}{\mu} \binom{n}{\nu} \frac{(q - 1)^{\mu+\nu}}{(q^m - 1)(q^n - 1)} \end{aligned}$$

Because not all weights between 1 and mn occur for rank 1 matrices, plots of actual probability densities show spikes and gaps. However, the plots of cumulative distributions are smoother and lead us to expect a limiting normal distribution as the size of the matrices goes to infinity. FIGURES 1 and 2 show this behavior quite well. (In order to plot the approximating normal distribution, we numerically computed the standard deviation of the weight distribution for the given m, n , and q .)

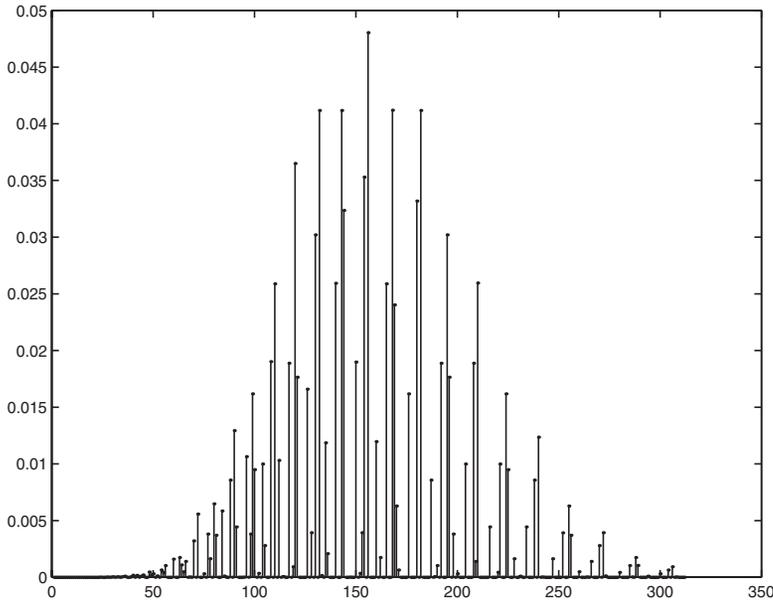


Figure 2 Density for the weight of rank 1 matrices, $m = n = 25, q = 2$

THEOREM 3. *As m or n goes to infinity, the weight distribution of rank 1 matrices approaches a normal distribution.*

Proof. The weight random variable for rank 1 matrices of size $m \times n$ is the product of independent binomial random variables conditioned on being positive. Define $W = XY$, where $X = \sum_{1 \leq i \leq m} X_i, Y = \sum_{1 \leq j \leq n} Y_j$, and X_i and Y_j are independent Bernoulli random variables with probability $1/q$ of being 0. Then W is the sum of m independent identically distributed random variables $X_i Y$. Conditioning W on $W > 0$ is the weight of rank 1 matrices. By the central limit theorem the distribution of W converges, as $m \rightarrow \infty$, to a normal distribution after suitable scaling. Now conditioning on W being positive does not change this result because the probability that $W > 0$ is $1 - q^{-m}$, which goes to 1 as $m \rightarrow \infty$. ■

Therefore, when m and n are large we can use a normal distribution of mean $\mathbf{E}(W)$ and variance $\text{var}(W)$ to approximate the weight distribution for rank 1 matrices. Note that this variance is not exactly the variance of the weight of rank 1 matrices because we have not conditioned on W being positive. However, the exact computation of that variance is rather complicated, and because of the theorem, the exact variance is not more informative than the variance of the unconditioned random variable W .

How fast does the variance grow as m and n go to infinity? For simplicity we let $m = n$, but the computation in the general case is similar. The variance of W is given by

$$\text{var}(W) = \mathbf{E}(W^2) - \mathbf{E}(W)^2.$$

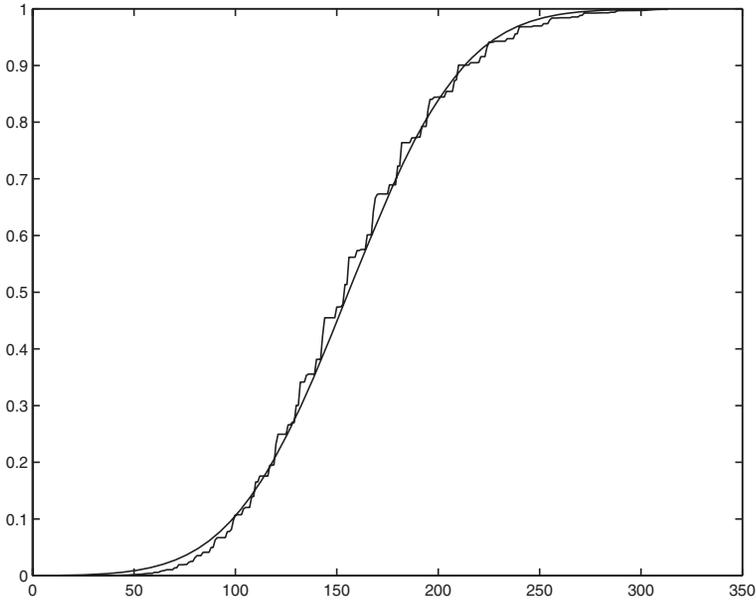


Figure 3 Cumulative frequency distribution for the weight of rank 1 matrices, $m = n = 25$, $q = 2$. The smooth curve is the normal distribution with the same mean ($\approx mn/4 = 156.25$) and standard deviation (≈ 44.63).

We need $\mathbf{E}(X)$ and $\mathbf{E}(X^2)$, which are given by

$$\mathbf{E}(X) = n(1 - 1/q)$$

$$\mathbf{E}(X^2) = n(1 - 1/q) + n(n - 1)(1 - 1/q)^2.$$

Because X and Y are independent binomial random variables with the same distribution, it follows that

$$\mathbf{E}(W) = \mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) = \mathbf{E}(X)^2$$

$$\mathbf{E}(W^2) = \mathbf{E}(X^2Y^2) = \mathbf{E}(X^2)\mathbf{E}(Y^2) = \mathbf{E}(X^2)^2.$$

After taking care of the algebra we arrive at

$$\text{var}(W) = \frac{1}{q^2} \left(1 - \frac{1}{q}\right)^2 n^2 + \frac{2}{q} \left(1 - \frac{1}{q}\right)^3 n^3.$$

From this we can see that for square matrices, the variance grows like n^3 , and the standard deviation grows like $n^{3/2}$.

Over the field with two elements, the variance is

$$\frac{n^3}{8} + \frac{n^2}{16},$$

and so the standard deviation is asymptotic to $(n/2)^{3/2}$. In the example shown in FIGURES 1 and 2, the standard deviation of the actual weight distribution of rank 1 matrices is 44.63 (rounded to two places). The value of $(n/2)^{3/2}$ with $n = 25$ is 44.19 (also rounded to two places).

Further questions

Analyzing the CR factorization for rank 2 matrices should conceivably allow us to find the weight distribution for rank 2, but the analysis is considerably more difficult, and for higher ranks the difficulty continues to increase. This suggests gathering some information by simulation. In FIGURE 3, we show a histogram for the weights of 10,000 matrices of rank 2 and size 25×25 over \mathbf{F}_2 . The average weight from Theorem 2 in this case is

$$\frac{25^2(1 - 1/2)(1 - 1/2)^2}{(1 - 1/2^{25})^2} = 234.3750 \dots$$

The sample mean is 234.4434, and the sample standard deviation is 39.5201. The histogram makes it plausible that the weight has a limiting normal distribution. In fact, for any k we expect a limiting normal distribution for the weight (with suitable scaling) of rank k matrices as the size goes to infinity.

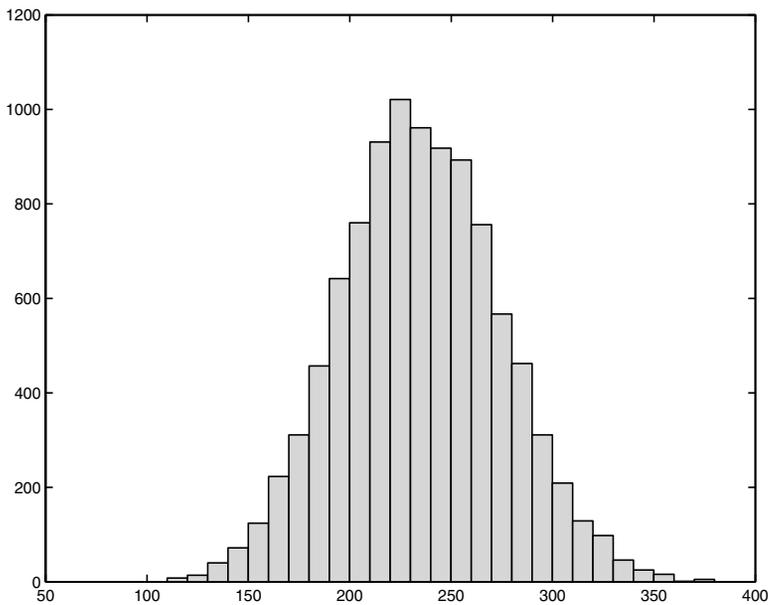


Figure 4 Histogram for the weight of 10,000 rank 2 matrices, $m = n = 25$, $q = 2$; bins have width 10

The question of simulation leads to the question of efficiently generating random matrices of fixed rank. Calabi and Wilf [1] treat the related problem of randomly generating a subspace of fixed dimension over a finite field, and Wilf has suggested to us that a random rank k matrix could be generated by adding together k matrices of rank 1, which are easy to generate, and then keeping those of rank k . Alternatively, one might use the CR factorization. Selecting R is exactly the subspace selection problem just mentioned. Selecting C can be done by generating a random $m \times k$ matrix and keeping those of rank k . Which approach is more efficient we leave as an open question, as well as the question of whether there are even better ways to generate matrices of a fixed rank.

We have focused on the weight of fixed rank matrices, but it would be interesting to look at the rank of fixed weight matrices. As an example, consider the $n \times n$ matrices of weight n . The number of these with rank 1 can be expressed in terms of the divisors

of n , using our previous result. Those of rank n are generalizations of permutation matrices and there are $n!(q-1)^n$ of them. What about the other ranks? In particular, how many $n \times n$ matrices of weight n and rank $n-1$ are there over \mathbf{F}_2 ?

Since the weight of A is the Hamming distance from A to 0 , it plays a role analogous to the norm of a real or complex matrix. In both cases it is the distance to the only matrix of rank 0 . Now we may ask for the distance from A to the subset of matrices of rank 1 , that is for the minimal distance from A to some matrix of rank 1 . In general we may ask for the distance from A to the matrices of rank k . For real or complex matrices these distances (using the linear map norm) are given by the *singular values* and can easily be computed [3, p. 468]. For matrices over finite fields can these distances (defined by the weight) be computed in any other way than by exhaustive search?

REFERENCES

1. E. Calabi and H. S. Wilf, On the sequential and random selection of subspaces over a finite field, *J. Combinatorial Theory (A)*, **22** (1977), 107–109; MR **55** #5649
2. H. Cohn, Projective geometry over \mathbf{F}_1 and the Gaussian binomial coefficients, *Amer. Math. Monthly*, **111** (2004), 487–495.
3. D. C. Lay, *Linear Algebra and Its Applications*, second edition, Addison-Wesley, Reading, Massachusetts, 1997.
4. J. P. S. Kung, The subset-subspace analogy, in *Gian-Carlo Rota on combinatorics*, 277–283, Birkhäuser, Boston, Boston, MA, 1995; MR 99b:01027
5. R. Lidl and H. Niederreiter, *Finite Fields*, second edition, Cambridge Univ. Press, Cambridge, UK, 1997; MR 97i:11115
6. W. P. Wardlaw, Row rank equals column rank, this MAGAZINE, **78** (2005), 316–318.

To appear in *The College Mathematics Journal* November 2006

Articles:

Playing Ball in a Space Station by *Andrew Simoson*

An Exceptional Exponential Function by *Branko Ćurgus*

More Designer Decimals: The Integers and Their Geometric Extensions by *O-Yeat Chan and James Smoak*

The Divergence of Balanced Harmonic-like Series by *Carl V. Lutzer and James E. Marengo*

An Interview with H.W. Gould by *Scott H. Brown*

Classroom Capsules

Another Look at Some p -Series by *Ethan Berkove*

Summing Cubes by Counting Rectangles by *Arthur T. Benjamin, Jennifer J. Quinn, and Calyssa Wurtz*

The Converse of Viviani's Theorem by *Zhibo Chen and Tian Liang*