

A note on the distribution of weights of fixed-rank matrices over the binary field

*Original*

A note on the distribution of weights of fixed-rank matrices over the binary field / Sanna, Carlo. - In: FINITE FIELDS AND THEIR APPLICATIONS. - ISSN 1071-5797. - STAMPA. - 87:(2023), p. 102157. [10.1016/j.ffa.2022.102157]

*Availability:*

This version is available at: 11583/2974495 since: 2023-01-11T07:48:42Z

*Publisher:*

Elsevier

*Published*

DOI:10.1016/j.ffa.2022.102157

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

Elsevier preprint/submitted version

Preprint (submitted version) of an article published in FINITE FIELDS AND THEIR APPLICATIONS © 2023,  
<http://doi.org/10.1016/j.ffa.2022.102157>

(Article begins on next page)

# A NOTE ON THE DISTRIBUTION OF WEIGHTS OF FIXED-RANK MATRICES OVER THE BINARY FIELD

CARLO SANNA<sup>†</sup>

ABSTRACT. Let  $\mathbf{M}$  be a random  $m \times n$  rank- $r$  matrix over the binary field  $\mathbb{F}_2$ , and let  $\text{wt}(\mathbf{M})$  be its Hamming weight, that is, the number of nonzero entries of  $\mathbf{M}$ .

We prove that, as  $m, n \rightarrow +\infty$  with  $r$  fixed and  $m/n$  tending to a constant, we have that

$$\frac{\text{wt}(\mathbf{M}) - \frac{1-2^{-r}}{2}mn}{\sqrt{\frac{2^{-r}(1-2^{-r})}{4}(m+n)mn}}$$

converges in distribution to a standard normal random variable.

## 1. INTRODUCTION

Let  $\mathbb{F}_q$  be a finite field of  $q$  elements and, for every matrix  $\mathbf{M}$  over  $\mathbb{F}_q$ , let  $\text{wt}(\mathbf{M})$  be the Hamming weight of  $\mathbf{M}$ , that is, the number of nonzero entries of  $\mathbf{M}$ .

Migler, Morrison, and Ogle [1] proved a formula for the expected value of  $\text{wt}(\mathbf{M})$  when  $\mathbf{M}$  is a random  $m \times n$  rank- $r$  matrix over  $\mathbb{F}_q$  taken with uniform probability. Moreover, they suggested that if  $m, n \rightarrow +\infty$ , with fixed  $r$  and  $q$ , then  $\text{wt}(\mathbf{M})$  approaches a normal distribution; and they made some considerations on the cases  $r = 1, 2$  (see Remark 1.1 below).

We prove the following result.

**Theorem 1.1.** *Fix a positive integer  $r$  and a real number  $\rho > 0$ . Let  $\mathbf{M}$  be a random  $m \times n$  rank- $r$  matrix over  $\mathbb{F}_2$  taken with uniform probability. Then, as  $m, n \rightarrow +\infty$  with  $m/n \rightarrow \rho$ , we have that*

$$\frac{\text{wt}(\mathbf{M}) - \frac{1-2^{-r}}{2}mn}{\sqrt{\frac{2^{-r}(1-2^{-r})}{4}(m+n)mn}}$$

converges in distribution to a standard normal random variable.

It might be interesting to strengthen Theorem 1.1 by letting also  $r$  goes to infinity, but sufficiently slowly in terms of  $m$  and  $n$ . Furthermore, one could consider analogs of Theorem 1.1 for matrices over an arbitrary finite field  $\mathbb{F}_q$ , or over rings such as  $\mathbb{Z}/n\mathbb{Z}$  (for a suitable definition of the rank). Then, instead of the Hamming weight, one could more generally consider the number of entries of  $\mathbf{M}$  that are equal to a prescribed fixed element.

*Remark 1.1.* Theorem 3 in [1] asserts that the weight distribution of  $m \times n$  rank-1 matrices over  $\mathbb{F}_q$  approaches a normal distribution as  $m, n \rightarrow +\infty$ . However, the proof provided in [1] is incorrect since, in order to apply the central limit theorem, it assumes that the random variables  $X_i Y$  are independent, while in fact they are not (they are all multiple of the same random variable  $Y$ ).

## 2. PRELIMINARIES

Hereafter, let  $m, n, r$  be positive integers with  $r \leq \min(m, n)$ . For every field  $\mathbb{K}$ , let  $\mathbb{K}^{m \times n}$  be the vector space of  $m \times n$  matrices with entries in  $\mathbb{K}$ , and let  $\mathbb{K}^{m \times n, r}$  be the set of matrices  $\mathbf{M} \in \mathbb{K}^{m \times n}$  such that  $\text{rank}(\mathbf{M}) = r$ .

---

2010 *Mathematics Subject Classification.* Primary: 15B52, 11T99 Secondary: 15B33, 05A16.

*Key words and phrases.* binary matrix; Hamming weight; normal distribution; random matrix; rank.

<sup>†</sup> C. Sanna is a member of GNSAGA of INdAM and of CryptTO, the group of Cryptography and Number Theory of Politecnico di Torino.

The following lemma regards the so-called “full rank factorization” of matrices and it is well known (cf. [2, Theorem 2]). We give a short proof for the sake of completeness.

**Lemma 2.1.** *Let  $\mathbb{K}$  be an arbitrary field. For every  $\mathbf{M} \in \mathbb{K}^{m \times n, r}$  there exist  $\mathbf{X} \in \mathbb{K}^{m \times r, r}$  and  $\mathbf{Y} \in \mathbb{K}^{r \times n, r}$  such that  $\mathbf{M} = \mathbf{X}\mathbf{Y}$ . Moreover, if  $\mathbf{M} = \mathbf{X}'\mathbf{Y}'$  for some  $\mathbf{X}' \in \mathbb{K}^{m \times r, r}$  and  $\mathbf{Y}' \in \mathbb{K}^{r \times n, r}$ , then there exists  $\mathbf{R} \in \mathbb{K}^{r \times r, r}$  such that  $\mathbf{X}' = \mathbf{X}\mathbf{R}$  and  $\mathbf{Y}' = \mathbf{R}^{-1}\mathbf{Y}$ .*

*Proof.* Pick  $\mathbf{X}$  has a matrix whose columns form a basis of the column space of  $\mathbf{M}$ . Note that indeed  $\mathbf{X} \in \mathbb{K}^{m \times r, r}$ . Since each column of  $\mathbf{M}$  can be uniquely written as a linear combination of the columns of  $\mathbf{X}$ , we get that  $\mathbf{M} = \mathbf{X}\mathbf{Y}$  for a unique  $\mathbf{Y} \in \mathbb{K}^{r \times n}$ . Therefore, we have that

$$\text{rank}(\mathbf{Y}) = \text{rank}(\mathbf{X}\mathbf{Y}) = \text{rank}(\mathbf{M}) = r,$$

and so  $\mathbf{Y} \in \mathbb{K}^{r \times n, r}$ . If  $\mathbf{M} = \mathbf{X}'\mathbf{Y}'$  for some  $\mathbf{X}' \in \mathbb{K}^{m \times r, r}$  and  $\mathbf{Y}' \in \mathbb{K}^{r \times n, r}$ , then the columns of  $\mathbf{X}'$  form a basis of the column space of  $\mathbf{M}$ . Hence, there exists  $\mathbf{R} \in \mathbb{K}^{r \times r, r}$  such that  $\mathbf{X}' = \mathbf{X}\mathbf{R}$ . Consequently, we have that

$$\mathbf{X}\mathbf{Y} = \mathbf{M} = \mathbf{X}'\mathbf{Y}' = \mathbf{X}\mathbf{R}\mathbf{Y}',$$

By the uniqueness of  $\mathbf{Y}$ , we get that  $\mathbf{Y} = \mathbf{R}\mathbf{Y}'$  and so  $\mathbf{Y}' = \mathbf{R}^{-1}\mathbf{Y}$ .  $\square$

We identify  $\mathbb{F}_2$  with  $\{0, 1\}$  and we let  $\oplus$  and  $\otimes$  denote the addition and multiplication of  $\mathbb{F}_2$ , respectively. The next lemma relates the operations of  $\mathbb{F}_2$  with the usual addition and multiplication of  $\mathbb{N}$ .

**Lemma 2.2.** *Let  $a_1, \dots, a_r \in \mathbb{F}_2$ . Then:*

- (i)  $\otimes_{k=1}^r a_k = \prod_{k=1}^r a_k$  and
- (ii)  $\oplus_{k=1}^r a_k = \sum_{\substack{S \subseteq \{1, \dots, r\} \\ S \neq \emptyset}} (-2)^{|S|-1} \prod_{k \in S} a_k$ .

*Proof.* Claim (i) is obvious. For claim (ii), let  $\mathcal{T} := \{k \in \{1, \dots, r\} : a_k = 1\}$ . Then

$$\bigoplus_{k=1}^r a_k = \begin{cases} 1 & \text{if } |\mathcal{T}| \text{ is odd} \\ 0 & \text{if } |\mathcal{T}| \text{ is even} \end{cases} = \frac{((1-2)^{|\mathcal{T}|} - 1)}{-2} = \sum_{\substack{S \subseteq \mathcal{T} \\ S \neq \emptyset}} (-2)^{|S|-1} = \sum_{\substack{S \subseteq \{1, \dots, r\} \\ S \neq \emptyset}} (-2)^{|S|-1} \prod_{k \in S} a_k,$$

as desired.  $\square$

In what follows, let  $\mathbf{X} \in \mathbb{F}_2^{m \times r}$  and  $\mathbf{Y} \in \mathbb{F}_2^{r \times n}$  be independent uniformly distributed random matrices. Moreover, for each  $\mathcal{S} \subseteq \{1, \dots, r\}$ , let

$$X_{\mathcal{S}} := \sum_{i=1}^m \prod_{k \in \mathcal{S}} x_{i,k} \quad \text{and} \quad Y_{\mathcal{S}} := \sum_{j=1}^n \prod_{k \in \mathcal{S}} y_{k,j},$$

and let also

$$Z := \sum_{i=1}^m \prod_{k=1}^r (1 - x_{i,k}) \quad \text{and} \quad W := \sum_{j=1}^n \prod_{k=1}^r (1 - y_{k,j}),$$

where  $x_{i,j}$  and  $y_{i,j}$  are the entries of  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively.

We shall need the following two lemmas.

**Lemma 2.3.** *We have*

$$\text{wt}(\mathbf{X}\mathbf{Y}) - \frac{1}{2}mn = \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} X_{\mathcal{S}} Y_{\mathcal{S}}.$$

*Proof.* By Lemma 2.2, we get that

$$\text{wt}(\mathbf{X}\mathbf{Y}) = \sum_{i=1}^m \sum_{j=1}^n \bigoplus_{k=1}^r (x_{i,k} \otimes y_{k,j}) = \sum_{i=1}^m \sum_{j=1}^n \bigoplus_{k=1}^r x_{i,k} y_{k,j} = \sum_{i=1}^m \sum_{j=1}^n \sum_{\substack{S \subseteq \{1, \dots, r\} \\ S \neq \emptyset}} (-2)^{|S|-1} \prod_{k \in S} x_{i,k} y_{k,j}.$$

Hence, since the empty product is equal to 1, we obtain that

$$\begin{aligned} \text{wt}(\mathbf{XY}) - \frac{1}{2}mn &= \sum_{i=1}^m \sum_{j=1}^n \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} \prod_{k \in \mathcal{S}} x_{i,k} y_{k,j} \\ &= \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} \sum_{i=1}^m \prod_{k \in \mathcal{S}} x_{i,k} \sum_{j=1}^n \prod_{k \in \mathcal{S}} y_{k,j} = \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} X_{\mathcal{S}} Y_{\mathcal{S}}, \end{aligned}$$

as claimed.  $\square$

**Lemma 2.4.** *We have*

$$\begin{aligned} \text{wt}(\mathbf{XY}) - \frac{1-2^{-r}}{2}mn &= \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} (X_{\mathcal{S}} - 2^{-|\mathcal{S}|}m)(Y_{\mathcal{S}} - 2^{-|\mathcal{S}|}n) \\ &\quad - \frac{1}{2}n(Z - 2^{-r}m) - \frac{1}{2}m(W - 2^{-r}n). \end{aligned}$$

*Proof.* From Lemma 2.3 and the identity

$$X_{\mathcal{S}} Y_{\mathcal{S}} = (X_{\mathcal{S}} - 2^{-|\mathcal{S}|}m)(Y_{\mathcal{S}} - 2^{-|\mathcal{S}|}n) + 2^{-|\mathcal{S}|}n X_{\mathcal{S}} + 2^{-|\mathcal{S}|}m Y_{\mathcal{S}} - 2^{-2|\mathcal{S}|}mn,$$

it follows that

$$\begin{aligned} (1) \quad \text{wt}(\mathbf{XY}) - \frac{1}{2}mn &= \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} (X_{\mathcal{S}} - 2^{-|\mathcal{S}|}m)(Y_{\mathcal{S}} - 2^{-|\mathcal{S}|}n) \\ &\quad - \frac{1}{2}n \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-1)^{|\mathcal{S}|} X_{\mathcal{S}} - \frac{1}{2}m \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-1)^{|\mathcal{S}|} Y_{\mathcal{S}} + \frac{1}{2}mn \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-\frac{1}{2})^{|\mathcal{S}|}. \end{aligned}$$

Furthermore, we have that

$$\begin{aligned} \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-1)^{|\mathcal{S}|} X_{\mathcal{S}} &= \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-1)^{|\mathcal{S}|} \sum_{i=1}^m \prod_{k \in \mathcal{S}} x_{i,k} \\ &= \sum_{i=1}^m \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} \prod_{k \in \mathcal{S}} (-x_{i,k}) = \sum_{i=1}^m \prod_{k=1}^r (1 - x_{i,k}) = Z, \end{aligned}$$

and similarly for the third sum in (1); while the fourth sum in (1) is equal to  $(1 - \frac{1}{2})^r = 2^{-r}$ .

Hence, we get that

$$\text{wt}(\mathbf{XY}) - \frac{1}{2}mn = \sum_{\mathcal{S} \subseteq \{1, \dots, r\}} (-2)^{|\mathcal{S}|-1} (X_{\mathcal{S}} - 2^{-|\mathcal{S}|}m)(Y_{\mathcal{S}} - 2^{-|\mathcal{S}|}n) - \frac{1}{2}nZ - \frac{1}{2}mW + \frac{1}{2}2^{-r}mn,$$

and the claim follows.  $\square$

**Lemma 2.5.** *We have that*

$$\mathbf{P}[\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{Y}) = r] \rightarrow 1,$$

as  $m, n \rightarrow +\infty$  with  $r$  fixed.

*Proof.* It is well-known (see, e.g., [1, Formula 3]) that

$$|\mathbb{F}_2^{m \times r, k}| = \prod_{i=0}^{k-1} \frac{(2^m - 2^i)(2^r - 2^i)}{2^k - 2^i},$$

for every nonnegative integer  $k \leq r$ . Therefore, we have that

$$\mathbf{P}[\text{rank}(\mathbf{X}) < r] = \frac{1}{2^{mr}} \sum_{k=0}^{r-1} |\mathbb{F}_2^{m \times r, k}| < \sum_{k=0}^{r-1} 2^{rk-m(r-k)} \rightarrow 0,$$

as  $m \rightarrow +\infty$  with  $r$  fixed. A similar reasoning gives that  $\mathbf{P}[\text{rank}(\mathbf{Y}) < r] \rightarrow 0$ , as  $n \rightarrow +\infty$  with  $r$  fixed. The claim follows.  $\square$

## 3. PROOF OF THEOREM 1.1

Fix a positive integer  $r$  and a real number  $\rho > 0$ , and assume that  $m, n \rightarrow +\infty$  with  $m/n \rightarrow \rho$ . By Lemma 2.5, the probability that  $\mathbf{X}$  and  $\mathbf{Y}$  have ranks equal to  $r$  tends to 1. Moreover, by Lemma 2.1, under the condition that  $\mathbf{X}$  and  $\mathbf{Y}$  have rank  $r$ , the random variable  $\mathbf{XY}$  is uniformly distributed in  $\mathbb{F}_2^{m \times n, r}$ . Therefore, for the sake of proving Theorem 1.1, we can assume that  $\mathbf{M} = \mathbf{XY}$ .

It can be easily checked that  $X_S$  and  $Y_S$  are binomial random variables of  $m$  and  $n$  trials, respectively, and probabilities of success equal to  $2^{-|S|}$ . Similarly,  $Z$  and  $W$  are binomial random variables of  $m$  and  $n$  trials, respectively, and probabilities of success equal to  $2^{-r}$ . For the sake of brevity, for each random variable  $T$  that has finite expected value and finite nonzero variance, we put  $T' := (T - \mathbf{E}[T])/\sqrt{\mathbf{V}[T]}$ . Then, by the central limit theorem, we have that  $X'_S, Y'_S, Z', W'$  converge in distribution to some standard normal random variables, which we call  $\hat{X}_S, \hat{Y}_S, \hat{Z}, \hat{W}$ , respectively.

Moreover, from Lemma 2.4, it follows that

$$(2) \quad \frac{\text{wt}(\mathbf{M}) - \frac{1-2^{-r}}{2}mn}{\sqrt{\frac{2^{-r}(1-2^{-r})}{4}(m+n)mn}} = \sum_{S \subseteq \{1, \dots, r\}} \frac{(-1)^{|S|-1}(1-2^{-|S|})}{\sqrt{2^{-r}(1-2^{-r})(m+n)}} X'_S Y'_S \\ - \frac{Z'}{\sqrt{1+m/n}} - \frac{W'}{\sqrt{1+n/m}}.$$

Since  $X'_S$  and  $Y'_S$  are independent, their product converges in distribution to the product  $\hat{X}_S \hat{Y}_S$ . Therefore, from Slutsky's theorem, we get that the sum in (2) converges in distribution to the constant 0. Consequently, the sum in (2) converges in probability to the constant 0.

Since  $Z'$  and  $W'$  are independent and  $m/n \rightarrow \rho$ , we get that

$$(3) \quad \frac{Z'}{\sqrt{1+m/n}} + \frac{W'}{\sqrt{1+n/m}} \rightarrow \frac{1}{\sqrt{1+\rho}} \hat{Z} + \frac{1}{\sqrt{1+1/\rho}} \hat{W}$$

in distribution. Also, since  $Z'$  and  $W'$  are independent, it follows that the right-hand-side of (3) is a standard normal random variable. Hence, again from Slutsky's theorem, we obtain that the left-hand-side of (2) converges in distribution to a standard normal random variable.

The proof is complete.

## REFERENCES

1. T. Migler, K. E. Morrison, and M. Ogle, *How much does a matrix of rank  $k$  weigh?*, Math. Mag. **79** (2006), no. 4, 262–271.
2. R. Piziak and P. L. Odell, *Full Rank Factorization of Matrices*, Math. Mag. **72** (1999), no. 3, 193–201.

DEPARTMENT OF MATHEMATICAL SCIENCES, POLITECNICO DI TORINO  
CORSO DUCA DEGLI ABRUZZI 24, 10129 TORINO, ITALY  
Email address: carlo.sanna@polito.it