A MODEL FOR RANDOM CHAIN COMPLEXES

MICHAEL J. CATANZARO AND MATTHEW ZABKA

Abstract. We introduce a random chain complex over a finite field. The randomness in our complex comes from choosing the entries in the matrices that represent the boundary maps uniformly over $\mathbb{F}_q$, conditioned on ensuring that the composition of consecutive boundary maps is the zero map. We then investigate the combinatorial and homological properties of this random chain complex.

1. Introduction

Topology's tools have become very popular for analyzing data, and as one usually regards data as a random vector, some have attempted to apply randomness to topological ideas. For example, Ginzburg and Pasechnik [1] have investigated a random chain complex with constant differential, while Zabka [4] has investigated a random Bockstein operator. Both of these papers have investigated their topics in a strictly algebraic setting, and in this paper, we too shall investigate a random chain complex in a strictly algebraic setting.

Chain complexes arise in topology as an algebraic measure in different dimensions of the relationship between the cycles and boundaries of a topological space. In particular, a chain complex defined on a space gives us a way to calculate that space's homology groups.

Formally, a *chain complex* (C_n, δ_n) is a sequence of modules, denoted C_n, and a sequence of linear transformations $\delta_n : C_n \to C_{n-1}$ that satisfy the 'boundary' condition $\delta_{n-1}\delta_n = 0$ for all n. The δ_n are usually called the boundary maps of the chain complex. An interested reader can see [2] for further details.

Let q be a prime number and let $\mathbb{F}_q$ denote the field with q elements. Consider the sequence of vector spaces $\mathbb{F}_q^{n_m}$ indexed by m in the integers. Let A_m be a random sequence of $(n_{m-1}) \times (n_m)$ matrices whose entries are chosen i.i.d. uniformly from $\mathbb{F}_q$, subject to the condition that the product of consecutive matrices is zero. We then consider the random chain complex $(\mathbb{F}_q^{n_m}, A_m)$. That is, we consider

$$\dots \xrightarrow{A_{m+1}} \mathbb{F}_q^{n_m} \xrightarrow{A_m} \mathbb{F}_q^{n_{m-1}} \xrightarrow{A_{m-1}} \dots$$

where $A_{i+1}A_i = 0$ for every i.

We have two main results. We first show that, as q goes to infinity, homology is concentrated in dimension zero. We then derive an explicit formula for the distribution of the Betti numbers.

Date: September 7, 2018.

2. Preliminaries

This section consists of several lemmas that count the number of elements in various sets related to finite vector spaces over $\mathbb{F}_q$. We provide proofs for these lemmas, but an interested reader can see [3] for further details.

Lemma 2.1. *The number of ordered, linearly independent k-tuples of vectors in $\mathbb{F}_q^n$ is*

$$(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{k-2})(q^n-q^{k-1}).$$

Proof. Since first vector in the k-tuple may be any vector except for the zero vector, there are q^n-1 choices for the first vector. For $1<m\le k$, the m-th vector in the k-tuple may be any vector that is not a linear combination of the previously chosen $m-1$ vectors. So there are q^n-q^{m-1} choices for the m-th vector. $\square$

The number $\begin{bmatrix} n \\ k \end{bmatrix}_q$ defined in the next theorem is known as the q-binomial coefficient.

Lemma 2.2. *The number of k-dimensional subspaces of $\mathbb{F}_q^n$ is*

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{k-2})(q^n-q^{k-1})}{(q^k-1)(q^k-q)(q^k-q^2)\cdots(q^k-q^{k-2})(q^k-q^{k-1})}.$$

Proof. Let $\begin{bmatrix} n \\ k \end{bmatrix}_q$ denote the number of k-dimensional subspaces of $\mathbb{F}_q^n$ and $N(q,k)$ be the number of ordered, linear independent k-tuples of vectors in $\mathbb{F}_q^n$. Then Corollary 2.1 gives

$$N(q,k)=(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{k-2})(q^n-q^{k-1}). \tag{1}$$

We may also find $N(q,k)$ another way: We can first choose a k-dimensional subspace and then choose the independent vectors in our k-tuple from the chosen subspace. There are $\begin{bmatrix} n \\ k \end{bmatrix}_q$ k-dimensional subspaces of $\mathbb{F}_q^n$. Then, there are q^k-1 choices for the first vector in the k-tuple, and, for $1<m\le k$, there are q^k-q^{m-1} vectors for the m-th vector in the k-tuple. Thus

$$N(q,k)=\begin{bmatrix} n \\ k \end{bmatrix}_q (q^k-1)(q^k-q)(q^k-q^2)\cdots(q^k-q^{k-2})(q^k-q^{k-1}). \tag{2}$$

Equations (1) and (2) give the required result. $\square$

Using Lemmas 2.2 and 1, we can count the number of matrices with of a given rank. The following lemma will aid us in Section 3.

Lemma 2.3. *The number of $m\times n$ matrices with entries in $\mathbb{F}_q$ of rank r is given by*

$$\begin{aligned}
&\begin{bmatrix} m \\ r \end{bmatrix}_q (q^n-1)(q^n-q)\cdots(q^n-q^{r-1})\\
=&\begin{bmatrix} n \\ r \end{bmatrix}_q (q^m-1)(q^m-q)\cdots(q^m-q^{r-1})\\
=&\frac{(q^m-1)(q^m-q)\cdots(q^m-q^{r-1})\cdot(q^n-1)(q^n-q)\cdots(q^n-q^{r-1})}{(q^r-1)(q^r-q)(q^r-q^2)\cdots(q^r-q^{r-2})(q^r-q^{r-1})}
\end{aligned}$$

Proof. Let W be a fixed r-dimensional subspace of $\mathbb{F}_q^n$. Then the number of matrices whose column space is W is given by the number of $r \times n$ matrices with rank r. This number is given by Lemma 2.1. The number of r-dimensional subspaces of $\mathbb{F}_q^n$ is $\begin{bmatrix} m \\ r \end{bmatrix}_q$, as stated in Lemma 2.2. The product of these is the number of $m \times n$ rank r matrices. □

We now turn our attention to a sequence of random matrices.**Matt:** I've moved this stuff to Section 2. We can talk about this

Definition 2.4. *Let n_m be a sequence of natural numbers. Let B_m be a sequence of random $(n_m) \times (n_{m-1})$ matrices whose entries are chosen i.i.d. uniformly from $\mathbb{F}_q$. Let r be a non-negative integer. Define*

$$P_k^m(r) := \mathbb{P}\left[\operatorname{rank}(B_{m+1}) = r | B_m B_{m+1} = 0, \operatorname{nul}(B_m) = k\right] .$$

Lemma 2.3 gives us the following.

Lemma 2.5. *With B_m defined as in Definition 2.4, we have that*

$$P_k^m(r) = \begin{cases} \dfrac{\left(\prod_{j=0}^{r-1}\left(q^{n_{m+1}} - q^j\right)\right)\left(\prod_{j=0}^{r-1}\left(q^k - q^j\right)\right)}{q^{kn_{m+1}}\left(\prod_{j=0}^{r-1}\left(q^r - q^j\right)\right)} & \text{if } k \neq 0, r \leq k \\ 0 & \text{if } r > k, \\ 1 & \text{if } r = k = 0. \end{cases}$$

Proof. Let $k = \operatorname{nul}(B_m)$ and suppose $B_m B_{m+1} = 0$. Then B_{m+1} maps $\mathbb{F}_q^{n_{m+1}}$ into the kernel of B_m, and thus $\operatorname{rank}(B_{m+1}) \leq k$. Therefore, if $r > k$, we have $P_k^m(r) = 0$. Further, if $k = 0$, then $\operatorname{rank}(B_{m+1}) = 0$, so $P_0^m(0) = 1$.

On the other hand, suppose $k \neq 0$ and $r \leq k$. Then, B_{m+1} represents a linear transformation from $\mathbb{F}_q^{n_{m+1}}$ into a k-dimensional subspace of $\mathbb{F}_q^{n_m}$. Thus, by changing basis, B_{m+1} can be represented by a $(n_{m+1}) \times k$ matrix. There are $q^{kn_{m+1}}$ such matrices, and, by Lemma 2.3, there are

$$\frac{\left(\prod_{j=0}^{r-1}\left(q^{n_{m+1}} - q^j\right)\right)\left(\prod_{j=0}^{r-1}\left(q^k - q^j\right)\right)}{\prod_{j=0}^{r-1}\left(q^r - q^j\right)}$$

such matrices of rank r. □

3. The Homology of a Random Chain Complex

Recall that a chain complex consists of a pair of sequences (C_n, δ_n), where the C_n are appropriate spaces (vector spaces, groups, modules, etc.) and the δ_n are maps, $\delta_n : C_n \to C_{n-1}$ such that $\delta_{n-1}\delta_n = 0$.

Let q be a prime number and let $\mathbb{F}_q$ denote the field with q elements. Let n_m be a sequence of natural numbers. Consider the sequence of finite vector spaces $\mathbb{F}_q^{n_m}$ indexed by m. We iteratively construct a sequence A_m of random $(n_{m-1}) \times (n_m)$ matrices.

Suppose we know A_m. Let A_{m+1} be a random $(n_m) \times (n_{m+1})$ matrix whose entries are chosen i.i.d. uniformly from $\mathbb{F}_q$, subject to the condition that $A_{m+1}A_m = 0$.

Definition 3.1. *Let n_m be a sequence of natural numbers. A* ***random chain complex*** *is a pair $(\mathbb{F}_q^{n_m}, A_m)$, where A_m is an iteratively defined sequence of random of $(n_{m-1}) \times (n_m)$ matrices, as defined above.*

We wish to investigate the probabilistic properties of the homology of a random chain complex. We are primarily interested in the distribution of the Betti numbers $\beta_m := \dim(H_m(A_*; \mathbb{F}_q^n))$. Recall that $H_m(A_*; \mathbb{F}_q^n) = \ker(A_m)/A_{m+1}(\mathbb{F}_q^n)$, so if $k = \ker(A_m)$, then $\beta_m = k - \operatorname{rank}(A_{m+1})$. We are therefore interested in the probabilistic properties of $\operatorname{rank}(A_{m+1})$ given that $A_m A_{m+1} = 0$ and that $\operatorname{nul}(A_m) = k$.

Remark 3.2. *As the A_m are uniformly distributed, Definition 2.4 immediately gives us that*

$$P_k^m(r) = \mathbb{P}\left[\beta_m = k - r | \operatorname{nul}(A_m) = k\right] .$$

Matt: Letting $k \le n_{m+1}$ seems fine to me, but what about Definition 2.4?

Theorem 3.3. *Let β_m be the m-th Betti number of a random chain complex. If $k \le n_{m+1}$, then*

$$\mathbb{P}\left[\beta_m = 0 | A_m A_{m+1} = 0, \operatorname{nul}(A_m) = k\right] \to 1 \text{ as } q \to \infty.$$

Proof. We have

$$\begin{aligned} & \mathbb{P}\left[\beta_m = 0 | A_m A_{m+1} = 0, \operatorname{nul}(A_m) = k\right] \\ = \; & P_k^m(k) \\ = \; & \frac{\prod_{j=0}^{k-1}(q^{n_{m+1}} - q^j)\prod_{j=0}^{k-1}(q^k - q^j)}{q^{kn_{m+1}}\prod_{j=0}^{k-1}(q^k - q^j)} \\ = \; & \frac{\prod_{j=0}^{k-1}(q^{n_{m+1}} - q^j)}{q^{kn_{m+1}}} \\ = \; & \prod_{j=0}^{k-1}(1 - q^{j - n_{m+1}}), \end{aligned}$$

which tends to 1 as $q \to \infty$. □

The previous theorem immediately leads to two corollaries.

Corollary 3.4. *Let b be a positive integer that is less than or equal to k. Then*

$$\mathbb{P}[\beta_m = b | A_m A_{m+1} = 0, \operatorname{nul}(A_m) = k] \to 0 \text{ as } q \to \infty.$$

Proof. We have that

$$\begin{aligned} & \mathbb{P}[\beta_m = b \mid A_m A_{m+1} = 0, \operatorname{nul}(A_m) = k] \\ = \; & 1 - \sum_{j \neq b} \mathbb{P}[\beta_0 = j \mid A_m A_{m+1} = 0, \operatorname{nul}(A_m) = k] \\ \le \; & 1 - \mathbb{P}(\beta_m = 0 | A_m A_{m+1} = 0, \dim\ker(A_m) = k). \end{aligned}$$

By Theorem 3.3, this goes to 0 as q goes to infinity. □

Corollary 3.5. $\mathbb{E}[\beta_m | A_m A_{m+1} = 0, \dim\ker(A_m) = k] \to n$ *as* $q \to \infty$.

Proof. The conditional expectation of the m-th Betti number is given by

$$\begin{aligned} & \mathbb{E}[\beta_m | A_m A_{m+1} = 0, \dim\ker(A_m) = k] \\ = & \sum_{b=1}^{n} b\mathbb{P}(\beta_m = b | A_m A_{m+1} = 0, \dim\ker(A_m) = k)\,. \end{aligned}$$

By Corollary 3.4, all terms with $b < n$ in this sum tend to 0 as q goes to infinity.

On the other hand, when $b = n$,

$$n\mathbb{P}(\beta_m = 0 | A_m A_{m+1} = 0, \dim\ker(A_m) = k),$$

which goes to n has q goes to infinity by Theorem 3.3. □

4. Conjectures

This section consists of a bunch of miscellaneous facts and conjectures that either we should include or not, but either way, I think we should understand them. In particular, I think these will be useful things to know to get a good grasp on what's really going on.

Conjecture 4.1. *Consider the following.*

1. *Fix r. As a function of k, $P_k^m(r)$ is decreasing on its support.*
2. *Fix k. As a function of r, $P_k^m(r)$ is increasing on its support.*
3. *$P_k^m(r)$ has a maximum when $k = r$, as either variable changes.*
4. *$P_{r+1}^m(r) > P_r^m(r-1)$.*

Mike: The following isn't a conjecture. We have a very simple proof by induction.

Theorem 4.2.

$$\mathbb{P}[\beta_j = b_j] = \sum_{i_j=0}^{n_j} P_{i_j}^j(i_j - b_j) \sum_{i_{j-1}=0}^{n_{j-1}} P_{i_{j-1}}^{j-1}(n_j - i_j) \cdots \sum_{i_1=0}^{n_1} P_{i_1}^1(n_2 - i_2) P_{n_0}^0(n_1 - i_1)$$

Question 4.3. *What is the max of $\mathbb{P}[\beta_1 = b]$, as a function of b? For what b is the max attained? The same questions for $\mathbb{P}[\beta_2 = b]$.*

Question 4.4. *What if $n_0 = w$ and $n > 0 = 2w$? Intuitively, we would expect these numbers to not depend on the degree of the chain complex. Something similar should be true for any sequence of (n_i) which converge.*

Question 4.5. *What does this formula say if there exists $N > 0$ so that $n_i = 0$ for all $i > N$? Specifically, $\mathbb{P}[\beta_i = 0] = 1$ and $\mathbb{P}[\beta_i > 0] = 0$ for $i > N$.*

Using the same logic as the previous theorem, I'm *pretty sure* we have the following.

Conjecture 4.6.

$$\mathbb{P}[\operatorname{rank} A_m = r_m] = \sum_{i_{m-1}=0}^{n_{m-1}} P_{i_{m-1}}^{m-1}(r_m) \sum_{i_{m-2}=0}^{n_{m-2}} P_{i_{m-2}}^{m-2}(n_{m-1} - i_{m-1}) \cdots \sum_{i_1=0}^{n_2} P_{i_1}^1(n_2 - i_2) P_{n_0}^0(n_1 - i_1)$$

Question 4.7. *What distribution does the random variable* $\operatorname{rank}(A_m)$ *follow? What are its moments?*

Conjecture 4.8. *Attempt to answer all of the following in cases: (i) n_i is eventually constant, (ii) $n_i \sim o(i)$.*

(1) $\mathbb{P}[\beta_m = 0] \to ?$ *as* $m \to \infty$.
(2) $\mathbb{P}[\beta_m = j] \to ?$ *as* $m \to \infty$, $j > 0$.
(3) $\mathbb{E}[\beta_m] = ?$ *as* $m \to \infty$.

References

[1] Viktor L Ginzburg and Dmitrii V Pasechnik. Random chain complexes. *Arnold Mathematical Journal*, pages 1–8, 2017.
[2] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2002.
[3] Richard P. Stanley. *Enumerative Combinatorics*, volume 1. Cambridge University Press, 2011.
[4] Matthew Zabka. A random bockstein operator. *Algebra and Discrete Mathematics*, 25(2):311–321, 2018.