Séminaire Lotharingien de Combinatoire XX (2018)
Article #YY, 12 pp.

Proceedings of the 30th Conference on Formal Power Series and Algebraic Combinatorics (Hanover)

Factorization Statistics and the Twisted Grothendieck-Lefschetz formula

Trevor Hyde*

Dept. of Mathematics
University of Michigan
Ann Arbor, MI 48109-1043

Abstract. Factorization statistics are functions defined on the set $\mathrm{Poly}_d(\mathbb{F}_q)$ of all monic degree d polynomials with coefficients in $\mathbb{F}_q$ which only depend on the degrees of the irreducible factors of a polynomial. We show that the expected values of factorization statistics are determined by the representation theoretic structure of the cohomology of configurations in $\mathbb{R}^3$. This *twisted Grothendieck-Lefschetz formula for* Poly_d is analogous to a result of Church, Ellenberg, and Farb for *squarefree* polynomials. Our result is shown with formal power series methods which also lead to a new proof of the Church, Ellenberg, and Farb result circumventing the machinery of algebraic geometry.

Keywords: arithmetic statistics, symmetric group representations, configuration space

1 Introduction

Arithmetic statistics is the study of arithmetically interesting rings like $\mathbb{Z}$ and $\mathbb{F}_q[x]$ from a probabilistic or statistical point of view. For example, one can ask for the probability that a random integer m in the interval $[1, n]$ is prime. The Prime Number Theorem tells us that

$$\mathrm{Prob}(m \in [1, n] \text{ is prime}) \approx \frac{1}{\log(n)}$$

for sufficiently large n. The analogous question for $\mathbb{F}_q[x]$ is to ask for the probability that a random monic degree d polynomial $f(x) \in \mathbb{F}_q[x]$ is irreducible. One can show that

$$\mathrm{Prob}(f(x) \text{ monic degree } d \text{ is irreducible}) \approx \frac{1}{d} \tag{1.1}$$

for large values of q. Note that the number of monic degree d polynomials in $\mathbb{F}_q[x]$ is q^d, hence $\frac{1}{d} = \frac{1}{\log_q(q^d)}$ parallels the result for $\mathbb{Z}$. Thus we see the beginning of a theme: analogous arithmetic statistical questions for $\mathbb{Z}$ and $\mathbb{F}_q[x]$ often have analogous answers. However, we can do

*tghyde@umich.edu

much better than this approximate answer to the irreducibility question in $\mathbb{F}_q[x]$. The number of irreducible monic degree d polynomials in $\mathbb{F}_q[x]$ is given by dth *necklace polynomial*,

$$\#\{f(x) \text{ irreducible monic degree } d\} = \frac{1}{d}\sum_{e|d}\mu(d/e)q^e,$$

where μ is the Möbius function. Therefore we have an exact determination of the irreducible probability,

$$\text{Prob}(f(x) \text{ monic degree } d \text{ is irreducible}) = \frac{1}{d}\sum_{e|d}\frac{\mu(d/e)}{q^{d-e}}. \tag{1.2}$$

There is a sense in analytic number theory that the leading terms in a probability like (1.1) are the important information—that the "error terms" are noise to be silenced. But if we look at the *exact* answers on the $\mathbb{F}_q[x]$ side of the analogy we see a different picture. Each term in (1.2) has a structural interpretation. For example, if $d = 6$, then

$$\text{Prob}(f(x) \text{ monic degree 6 is irreducible}) = \frac{1}{6}\left(1 - \frac{1}{q^3} - \frac{1}{q^4} + \frac{1}{q^5}\right).$$

The four terms in this expression correspond to the intermediate fields of the degree 6 extension $\mathbb{F}_{q^6}/\mathbb{F}_q$; the coefficients encode how these fields fit together—far from noise!

This brings us to our main thesis: the exact expressions for arithmetic statistical questions in $\mathbb{F}_q[x]$ reflect hidden structure which is not apparent from approximations. In other words, *there are no error terms*, each term has an interpretation and together they tell a complete story. Our main result exemplifies this perspective.

Theorem 1.1 (Twisted Grothendieck-Lefschetz for Poly_d)**.** *Let P be a function defined on the set $\text{Poly}_d(\mathbb{F}_q)$ of monic degree d polynomials in $\mathbb{F}_q[x]$ such that $P(f)$ only depends on the degrees of the irreducible factors of f. P may also be viewed as a function defined on partitions of d, or equivalently as a class function of the symmetric group S_d. Let ψ_d^k be the character of the S_d-representation $H^{2k}(\text{PConf}_d(\mathbb{R}^3),\mathbb{Q})$ where $\text{PConf}_d(\mathbb{R}^3)$ is the ordered configuration space of d distinct points in $\mathbb{R}^3$ (see Section 3.) Then we can express the expected value $E_d(P)$ of P on $\text{Poly}_d(\mathbb{F}_q)$ as*

$$E_d(P) := \frac{1}{q^d}\sum_{f\in\text{Poly}_d(\mathbb{F}_q)} P(f) = \sum_{k=0}^{d-1}\frac{\langle P, \psi_d^k\rangle}{q^k},$$

where $\langle P, \psi_d^k\rangle = \frac{1}{d!}\sum_{\sigma\in S_d} P(\sigma)\psi_d^k(\sigma)$ is the standard inner product of class functions of the symmetric group S_d.

Theorem 1.1 gives a representation theoretic interpretation for the expected values of *factorization statistics* P. In Section 2 we give a precise definition of factorization statistics and survey examples to get a better feel for this family of arithmetic statistical questions. The twisted Grothendieck-Lefschetz formula provides a bridge to translate back and forth between representation theory and topology on the one hand and the combinatorics of finite fields on the other. We explore this interplay with some interesting examples in Section 4.

2 Factorization statistics

Let $\mathbb{F}_q$ be a finite field. The *factorization type* of a polynomial $f(x) \in \mathbb{F}_q[x]$ is the partition of $\deg(f)$ given by the degrees of the irreducible factors of $f(x)$. Let $\mathrm{Poly}_d(\mathbb{F}_q)$ denote the set of monic degree d polynomials in $\mathbb{F}_q[x]$. A *factorization statistic* P is a function defined on polynomials $f(x) \in \mathrm{Poly}_d(\mathbb{F}_q)$ which only depends on the factorization type of $f(x)$.

Example 2.1. We illustrate these notions with some examples.

1. Consider the polynomials $g(x), h(x) \in \mathrm{Poly}_5(\mathbb{F}_3)$ with irreducible factorizations

$$g(x) = x^2(x+1)(x^2+1) \qquad h(x) = (x+1)(x-1)(x^3-x+1).$$

   The factorization type of $g(x)$ is the partition $[2,1,1,1]$ and the factorization type of $h(x)$ is $[3,1,1]$. Note that the factorization type does not detect the multiplicity of a specific factor so that x^2 and $x(x+1)$ both have the same factorization type $[1,1]$.

2. Let $R(f)$ be the number of $\mathbb{F}_q$-roots of $f(x) \in \mathrm{Poly}_d(\mathbb{F}_q)$. Then $R(f)$ depends only on the number of linear factors of $f(x)$, hence is a factorization statistic. Referring to the two polynomials above we have $R(g) = 3$ and $R(h) = 2$.

3. Say a polynomial $f(x)$ has *even type* if the factorization type of $f(x)$ is an even partition. More specifically, say $\lambda = (1^{m_1}2^{m_2}3^{m_3}\cdots)$ is the factorization type of $f(x)$ and define $\mathrm{sgn}(\lambda)$ by

$$\mathrm{sgn}(\lambda) = \prod_{j\geq 1}(-1)^{m_j(j-1)},$$

   then $f(x)$ has even type if $\mathrm{sgn}(\lambda) = 1$. The function ET defined as

$$ET(f) = \begin{cases} 1 & f(x) \text{ has even type} \\ 0 & \text{otherwise,} \end{cases}$$

   is a factorization statistic. Continuing our examples, $ET(g) = 0$ and $ET(h) = 1$.

4. Define the *quadratic excess* $Q(f)$ of a polynomial to be

$$Q(f) = \#\{\text{reducible quadratic factors of } f(x)\} - \#\{\text{irreducible quadratic factors of } f(x)\}.$$

   Then $Q(f)$ depends only on the number of linear and irreducible quadratic factors of $f(x)$, hence is a factorization statistic. Since $g(x)$ has 3 linear factors and 1 irreducible quadratic factor, we have

$$Q(g) = \binom{3}{2} - 1 = 2.$$

   The polynomial $h(x)$ has 2 linear factors and no irreducible quadratic factors, hence

$$Q(h) = \binom{2}{2} - 0 = 1.$$

Say P is a factorization statistic, then we write

$$E_d(P) := \frac{1}{q^d} \sum_{f \in \mathrm{Poly}_d(\mathbb{F}_q)} P(f)$$

for the expected value of P on the set $\mathrm{Poly}_d(\mathbb{F}_q)$ of all monic degree d polynomials. By counting the number of polynomials with a given factorization type we can explicitly compute $E_d(P)$ for any particular P and d as a function of q. For example, here are some computations of $E_d(Q)$ where Q is the quadratic excess statistic defined in Example 2.1 (4).

d	$E_d(Q)$
3	$\frac{2}{q} + \frac{1}{q^2}$
4	$\frac{2}{q} + \frac{2}{q^2} + \frac{2}{q^3}$
5	$\frac{2}{q} + \frac{2}{q^2} + \frac{4}{q^3} + \frac{2}{q^4}$
6	$\frac{2}{q} + \frac{2}{q^2} + \frac{4}{q^3} + \frac{4}{q^4} + \frac{3}{q^5}$
10	$\frac{2}{q} + \frac{2}{q^2} + \frac{4}{q^3} + \frac{4}{q^4} + \frac{6}{q^5} + \frac{6}{q^6} + \frac{8}{q^7} + \frac{8}{q^8} + \frac{5}{q^9}$

There are a few remarkable features to these expected values. In each case $E_d(Q)$ is a polynomial in $\frac{1}{q}$ of degree $d-1$ with *positive integer coefficients*—one should expect the coefficients to be rational numbers, but both the positivity and integrality are not a priori evident. Moreover, if we evaluate the polynomial $E_d(Q)$ at $q = 1$ we get the binomial coefficient $\binom{d}{2}$ in each case. Note that the coefficients of $E_d(Q)$ appear to stabilize as d increases with a clear pattern emerging already for $d = 10$, suggesting that the expected values $E_d(Q)$ converge coefficientwise as $d \to \infty$.

It is also interesting to consider the expected value of factorization statistics on subsets of $\mathrm{Poly}_d(\mathbb{F}_q)$. A nice example is the subset $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ of squarefree polynomials. If P is a factorization statistic, then we write

$$E_d^{\mathrm{sf}}(P) := \frac{1}{|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)|} \sum_{f \in \mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)} P(f)$$

for the expected value of P on $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$. Recall that

$$|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)| = \begin{cases} q^d & d = 0, 1 \\ q^d - q^{d-1} & d \geq 2. \end{cases}$$

Returning to the quadratic excess example we compute the following table for $E_d^{\mathrm{sf}}(Q)$.

d	$E_d^{\mathrm{sf}}(Q)$
3	$-\frac{1}{q}$
4	$-\frac{1}{q}+\frac{2}{q^2}$
5	$-\frac{1}{q}+\frac{3}{q^2}-\frac{2}{q^3}$
6	$-\frac{1}{q}+\frac{3}{q^2}-\frac{4}{q^3}+\frac{2}{q^4}$
10	$-\frac{1}{q}+\frac{3}{q^2}-\frac{4}{q^3}+\frac{4}{q^4}-\frac{5}{q^5}+\frac{7}{q^6}-\frac{8}{q^7}+\frac{4}{q^8}$

The results are similar but with a few notable differences: $E_d^{\mathrm{sf}}(Q)$ is a degree $d-2$ polynomial in $\frac{1}{q}$ with alternating integral coefficients. Evaluating $E_d^{\mathrm{sf}}(Q)$ at $q=-1$ gives $\binom{d-1}{2}$. We still see coefficientwise stability as d increases, although the limit is less transparent.

These observations suggest there is some hidden structure underlying the expected values of factorization statistics. In the next section we explain how all of this relates to the topology of configuration spaces and the representation theory of the symmetric group.

3 Twisted Grothendieck-Lefschetz formulas

We briefly detour from our discussion of factorization statistics and finite fields to review some topology. Let X be any topological space, then $\mathrm{PConf}_d(X)$ is

$$\mathrm{PConf}_d(X)=\{(x_1,x_2,\dots,x_d)\in X^d : x_i\neq x_j\},$$

the *ordered configuration space of d points on* X. The symmetric group S_d acts on $\mathrm{PConf}_d(X)$ by permuting coordinates, and this action is free given our stipulation that all coordinates be different. Let $\mathrm{Conf}_d(X)$ be the quotient of $\mathrm{PConf}_d(X)$ by this action, then $\mathrm{Conf}_d(X)$ may be viewed as the *unordered configuration space of d points on* X. Note that for each $k\geq 0$ the singular cohomology $H^k(\mathrm{PConf}_d(X),\mathbb{Q})$ is, by functoriality, a finite dimensional S_d-representation.

If $X=\mathbb{C}$ is the complex plane, then there is a natural correspondence between $\mathrm{Conf}_d(\mathbb{C})$ and squarefree monic degree d polynomials $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{C})$ given by

$$\{\alpha_1,\alpha_2,\dots,\alpha_d\}\quad\longleftrightarrow\quad(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_d)$$

This correspondence gives us a conceptual link between configuration space and squarefree polynomials which helps us interpret the following result of Church, Ellenberg, and Farb [3, Prop. 4.1]. A couple remarks before we state the theorem:

1. The unit group $\mathbb{C}^\times$ acts on $\mathrm{PConf}_d(\mathbb{C})$ by simultaneously scaling all coordinates; this action commutes with S_d, hence $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times,\mathbb{Q})$ is also an S_d-representation.

2. If P is a factorization statistic defined on $\mathrm{Poly}_d(\mathbb{F}_q)$ then we may also view P as a function of the partitions of d, or equivalently as a class function defined on the symmetric group S_d.

Theorem 3.1 (Twisted Grothendieck-Lefschetz formula for $\mathrm{Poly}_d^{\mathrm{sf}}$)**.** *Let P be a factorization statistic, and let χ_d^k be the character of the S_d-representation $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$. Then the expected value $E_d^{\mathrm{sf}}(P)$ of P on the set $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ of squarefree polynomials is given by*

$$E_d^{\mathrm{sf}}(P) = \sum_{k=0}^{d-2} \frac{(-1)^k \langle P, \chi_d^k \rangle}{q^k},$$

where $\langle P, \chi_d^k \rangle = \frac{1}{d!} \sum_{\sigma \in S_d} P(\sigma)\chi_d^k(\sigma)$ is the standard inner product of class functions on the symmetric group S_d.

Theorem 3.1 expresses the coefficients of $E_d^{\mathrm{sf}}(P)$ in terms of the S_d-representation structure of the cohomology of $\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times$. Note that the factorization statistic P plays different roles on each side of this equation: on the left it acts as a function on the set $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ of squarefree polynomials; on the right it acts as a class function of the symmetric group S_d.

Church, Ellenberg, and Farb [3, Prop. 4.1] proved a variant of Theorem 3.1. The equivalence of their result and Theorem 3.1 is a consequence of work by the author and Jeff Lagarias [9, Thm. 4.3]. Their approach uses algebraic geometry: viewing PConf_d as a scheme defined over $\mathbb{Z}$, the Grothendieck-Lefschetz trace formula for étale cohomology with “twisted coefficients” expresses the weighted point counts on $\mathrm{Conf}_d(\mathbb{F}_q) \cong \mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ in terms of the trace of Frobenius. This combined with a purity result and a comparison theorem between étale and singular cohomology yields the twisted Grothendieck-Lefschetz formula for $\mathrm{Poly}_d^{\mathrm{sf}}$.

Recently the author [7] found a new proof of Theorem 3.1 using formal power series methods—entirely circumventing the machinery of algebraic geometry. Combinatorial descriptions of the cohomology of $\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times$ lead to beautiful product formulas for the *cycle index series* encoding the entire family of symmetric group representations; expanding these product formulas leads to an interpretation of the cycle index series as a generating function for the *squarefree splitting measure* which computes the probability of a squarefree polynomial over $\mathbb{F}_q$ having a given factorization type.

Let us return to the quadratic excess factorization statistic Q and see how Theorem 3.1 explains our observations about $E_d^{\mathrm{sf}}(Q)$. Recall that $Q(f)$ is defined to be the difference between the number of reducible versus irreducible quadratic factors of f. Rephrasing this in terms of partitions, let $x_k(\lambda)$ be the number of parts of λ of size k, then

$$Q(\lambda) = \binom{x_1(\lambda)}{2} - \binom{x_2(\lambda)}{1}.$$

Next let $\mathbb{Q}[d]$ be the permutation representation of the symmetric group with basis $\{e_1, e_2, \ldots, e_d\}$ and consider the representation given by the second exterior power $\bigwedge^2 \mathbb{Q}[d]$. This representation has dimension $\binom{d}{2}$ with basis given by $\{e_i \wedge e_j : i < j\}$. If $\sigma \in S_d$ is a permutation, then the trace of σ on $\bigwedge^2 \mathbb{Q}[d]$ is

$$\begin{aligned}\text{Trace}(\sigma) &= \#\{\{i,j\} : \sigma \text{ fixes } i \text{ and } j\} - \#\{\{i,j\} : \sigma \text{ transposes } i \text{ and } j\} \\ &= \binom{x_1(\sigma)}{2} - \binom{x_2(\sigma)}{1} \\ &= Q(\sigma).\end{aligned}$$

Thus Q, viewed as a class function of S_d, is the character of $\bigwedge^2 \mathbb{Q}[d]$. It follows that $\langle Q, \chi_d^k \rangle$ is a non-negative integer for all $d, k \geq 0$. This together with Theorem 3.1 explains the alternating integral coefficients of $E_d^{\text{sf}}(Q)$. That the degree of $E_d^{\text{sf}}(Q)$ is $d-2$ is a reflection of $d-2$ being the largest non-vanishing degree of cohomology for $\text{PConf}_d(\mathbb{C})/\mathbb{C}^\times$.

The coefficientwise convergence of $E_d^{\text{sf}}(Q)$ is a consequence of the *representation stability* of $H^k(\text{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ for each k. Representation stability is a phenomenon introduced by Church and Farb [4] to describe, for example, sequences A_d of S_d-representations whose decomposition into irreducible S_d-modules stabilizes in a technical sense as $d \to \infty$. See [3, 4] for more details on the theory of representation stability. The key point for this discussion is that the representation stability of $H^k(\text{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ implies that $\langle Q, \chi_d^k \rangle$ is eventually constant as $d \to \infty$. In [9, Prop. 6.2] the author and Lagarias show $H^k(\text{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ is isomorphic to the *rank-selected homology of the partition lattice* $\beta_{[k-1]}(\Pi_d)$ as S_d-representations. Hersh and Reiner [5, Thm. 10.1] proved that for each k the $\beta_{[k-1]}(\Pi_d)$ exhibit representation stability.

The twisted Grothendieck-Lefschetz formula for $\text{Poly}_d^{\text{sf}}$ interprets and explains the structure observed in the table of values for $E_d^{\text{sf}}(Q)$ in terms of the cohomology of configurations in $\mathbb{C}$ and the representation theory of the symmetric group. However, we made analogous observations for the expected values $E_d(Q)$ of the quadratic excess on the set $\text{Poly}_d(\mathbb{F}_q)$ of all monic degree d polynomials and yet Theorem 3.1 tells us nothing in this case. Thus it is natural to ask whether there is an analog of the twisted Grothendieck-Lefschetz formula for Poly_d.

We answer affirmatively with our main result.

Theorem 3.2 (Twisted Grothendieck-Lefschetz formula for Poly_d)**.** *Let P be a factorization statistic and let ψ_d^k be the character of the S_d-representation $H^{2k}(\text{PConf}_d(\mathbb{R}^3), \mathbb{Q})$. Then the expected value $E_d(P)$ of P on the set $\text{Poly}_d(\mathbb{F}_q)$ of polynomials is given by*

$$E_d(P) = \sum_{k=0}^{d-1} \frac{\langle P, \psi_d^k \rangle}{q^k}.$$

Theorem 3.2 is similar to Theorem 3.1 but with a few important differences. There is no alternating sign in this expression for $E_d(P)$ and the highest degree is $d-1$ instead of $d-2$. Most importantly there is a different family of representations determining the coefficients. We prove this result using formal power series methods parallel to the strategy described for our new proof of Theorem 3.1. Unlike the squarefree case, there is no apparent conceptual link between the set of all monic degree d polynomials $\mathrm{Poly}_d(\mathbb{F}_q)$ and the configuration space $\mathrm{PConf}_d(\mathbb{R}^3)$. This makes it difficult to envision how an algebro-geometric proof of Theorem 3.2 might proceed.

The family of representations $A_d^k := H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ appears in the literature under many guises, including as the higher Lie representations and as sign twists of the Eulerian idempotent projections of the regular representation $\mathbb{Q}[S_d]$. A survey of these various interpretations will appear in forthcoming work of the author and Lagarias [8]. Thus it may be that the conceptually correct interpretation of the A_d^k comes from another perspective, but given the nice geometric story in the squarefree case it seems reasonable to expect something analogous for $\mathrm{Poly}_d(\mathbb{F}_q)$.

Returning to our quadratic excess example, the observation that Q is the character of $\bigwedge^2 \mathbb{Q}[d]$ and Theorem 3.2 together immediately explain the positive integral coefficients noted in the table for $E_d(Q)$. Furthermore, it is known [2, Thm. 1] that for each k the sequence $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ of representations exhibits representation stability, hence the coefficientwise convergence of $E_d(Q)$ as $d \to \infty$.

The coefficientwise convergence of $E_d(P)$ and $E_d^{\mathrm{sf}}(P)$ also follows combinatorially for a large class of factorization statistics P called *character polynomials* of which Q is an example. This convergence phenomenon for expected values of factorization statistics holds in much greater generality, even when there is no representation stability present; we do not pursue this further here but refer the reader to [1, Cor. 10], [7]. One benefit of the combinatorial approach is that we can explicitly compute the limits of $E_d(P)$ and $E_d^{\mathrm{sf}}(P)$, finding them to be given by rational functions of q. For instance, in [7] we show

$$\lim_{d\to\infty} E_d(Q) = \frac{1}{2}\left(1+\frac{1}{q}\right)\left(\frac{1}{1-\frac{1}{q}}\right)^2 - \frac{1}{2}\left(1-\frac{1}{q}\right)\left(\frac{1}{1-\frac{1}{q^2}}\right)$$

$$\lim_{d\to\infty} E_d^{\mathrm{sf}}(Q) = \frac{1}{2}\left(1-\frac{1}{q}\right)\left(\frac{1}{1+\frac{1}{q}}\right)^2 - \frac{1}{2}\left(1-\frac{1}{q}\right)\left(\frac{1}{1+\frac{1}{q^2}}\right).$$

4 Examples

The twisted Grothendieck-Lefschetz formulas form a bridge between polynomial factorization statistics on the one hand and symmetric group representation theory and cohomology of configuration spaces on the other. Translating information back and forth across this bridge leads to an interesting interplay among these structures.

To illustrate we begin with the following constraint on the cohomology of configuration space.

Theorem 4.1. *For each $d \geq 0$ there is an isomorphism of S_d-representations*

$$\bigoplus_{k=0}^{d-1} H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}) \cong \mathbb{Q}[S_d], \tag{4.1}$$

where $\mathbb{Q}[S_d]$ is the regular representation of S_d.

Theorem 4.1 follows from Theorem 3.2 by considering the $q = 1$ evaluation of the *general splitting measure*. See [8] for details.

The right hand side of this isomorphism is well-understood: the irreducible representations of S_d are indexed by partitions $\lambda \vdash d$, each irreducible $\mathcal{S}_\lambda$ is a direct summand of $\mathbb{Q}[S_d]$ with multiplicity $f_\lambda := \dim \mathcal{S}_\lambda$. Thus Theorem 4.1 tells us that all the irreducible components of $\mathbb{Q}[S_d]$ are distributed among the various degrees of cohomology on the left hand side of (4.1). Theorem 3.2 implies that this filtration of the regular representation completely determines and is determined by the expected values of factorization statistics on $\mathrm{Poly}_d(\mathbb{F}_q)$. We use this information to locate some of the irreducible S_d-representations in the cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$.

4.1 Trivial representation

Let $\mathbf{1} = \mathcal{S}_{[d]}$ be the trivial representation of S_d. Recall that the trivial representation $\mathbf{1}$ is one dimensional with constant character equal to 1. By Theorem 4.1 there is precisely one k such that $\mathbf{1}$ is a summand of $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$. Interpreting the character of $\mathbf{1}$ as a factorization statistic we have $E_d(1) = 1$ and Theorem 3.2 implies

$$1 = E_d(1) = \sum_{k=0}^{d-1} \frac{\langle 1, \psi_d^k \rangle}{q^k}.$$

Comparing coefficients of $1/q^k$ we conclude that $\mathbf{1}$ is a summand of $H^0(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$. On the other hand, $\mathrm{PConf}_d(\mathbb{R}^3)$ is path connected so the degree 0 cohomology is one dimensional. Thus

$$H^0(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}) \cong \mathbf{1}.$$

Note that any factorization statistic P is a class function of S_d and the irreducible characters of S_d form a $\mathbb{Q}$-basis for the vector space of all class functions. Thus there are $a_\lambda \in \mathbb{Q}$ such that

$$P = \sum_{\lambda \vdash d} a_\lambda \chi_\lambda,$$

where χ_λ is the character of the irreducible representation $\mathcal{S}_\lambda$. In particular if $a_1 := a_{[d]}$ is the coefficient of the trivial character in this decomposition, then we have the following corollary.

Corollary 4.2. *If P is any factorization statistic and a_1 is the coefficient of the trivial character in the canonical expression for P as a linear combination of irreducible characters, then*

$$a_1 = \lim_{q \to \infty} E_d(P).$$

Hence $a_1 = 0$ if and only if the expected value of P approaches 0 for large q.

Our table of values for $E_d(Q)$ with Q the quadratic excess show that $\lim_{q\to\infty} E_d(Q) = 0$ for each d, hence the representation $\bigwedge^2 \mathbb{Q}[d]$ has no trivial component.

4.2 Sign representation

The only other one dimensional irreducible representation of S_d is the sign representation $\mathbf{Sgn} := \mathcal{S}_{[1^d]}$ whose character we write as sgn. Viewing sgn as a factorization statistic Theorem 3.2 implies

$$E_d(\mathrm{sgn}) = \frac{1}{q^k}$$

for some $k > 0$, but which value of k is it?

Theorem 4.3. *For each $d \geq 0$ we have*

$$E_d(\mathrm{sgn}) = \frac{1}{q^{\lfloor d/2 \rfloor}}.$$

Hence $H^{2\lfloor d/2 \rfloor}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ is the unique cohomological degree with a **Sgn** *summand.*

We prove Theorem 4.3 using our theory of *liminal reciprocity* which relates factorization statistics in $\mathrm{Poly}_d(\mathbb{F}_q)$ with the limiting values of *squarefree* factorization statistics for $\mathbb{F}_q[x_1, x_2, \ldots, x_n]$ as the number of variables n tends to infinity. See our forthcoming paper [6] for details.

Theorem 4.3 has a surprising consequence. Recall that ET is the *even type* factorization statistic defined as $ET(f) = 1$ when the factorization type of f is an even partition and 0 otherwise. Thus the expected value $E_d(ET)$ is the probability of a random polynomial in $\mathrm{Poly}_d(\mathbb{F}_q)$ having even factorization type. Off the cuff one might guess that a polynomial should be just as likely to have an even versus odd factorization type. However, notice that

$$ET = \tfrac{1}{2}(1 + \mathrm{sgn})$$

as class functions of S_d. It follows by the linearity of expectation that

$$E_d(ET) = \tfrac{1}{2}(E_d(1) + E_d(\mathrm{sgn})) = \tfrac{1}{2}\left(1 + \tfrac{1}{q^{\lfloor d/2 \rfloor}}\right).$$

The leading term of this probability is $1/2$ as we expected, but there is a bias toward a polynomial having even factorization type coming from the sign representation and the degree of cohomology in which it appears. For comparison we remark that in the squarefree case the probability of a random polynomial in $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ having even factorization type is exactly

$$E_d^{\mathrm{sf}}(ET) = \tfrac{1}{2},$$

matching our original guess. This is equivalent by Theorem 3.1 to $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ having no **Sgn** component for any k.

4.3 Standard representation

Recall the factorization statistic R from Example 2.1 where $R(f)$ is the number of $\mathbb{F}_q$-roots of $f(x)$. In [7] we use a formal power series argument to compute the expected number of roots $E_d(R)$ of a degree d polynomial to be

$$E_d(R) = \frac{1-\frac{1}{q^d}}{1-\frac{1}{q}} = 1 + \frac{1}{q} + \frac{1}{q^2} + \frac{1}{q^3} + \ldots + \frac{1}{q^{d-1}}. \tag{4.2}$$

Viewed as a class function of S_d, $R(\sigma)$ is the number of fixed points of σ. Hence R is the character of the permutation representation $\mathbb{Q}[d]$. It is well known that the irreducible decomposition of $\mathbb{Q}[d]$ is

$$\mathbb{Q}[d] \cong \mathbf{1} \oplus \mathbf{Std},$$

where $\mathbf{Std} := \mathcal{S}_{[d-1,1]}$ is the *standard representation* of S_d of dimension $d-1$. We already determined that $H^0(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}) \cong \mathbf{1}$, explaining the constant term in (4.2). Thus Theorem 3.2 implies that each $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ has a single $\mathbf{Std}$ component for $1 \leq k \leq d-1$, accounting for all copies of $\mathbf{Std}$. For comparison we note that

$$E_d^{\mathrm{sf}}(R) = \frac{1-\frac{(-1)^{d-1}}{q^{d-1}}}{1+\frac{1}{q}} = 1 - \frac{1}{q} + \frac{1}{q^2} - \frac{1}{q^3} + \ldots + (-1)^{d-1}\frac{1}{q^{d-2}}.$$

What does this tell us about $\mathbf{Std}$ components in $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$?

4.4 Evaluating at $q = 1$

Recall that the inner product $\langle \chi, \psi \rangle$ of symmetric group class functions is bilinear. If P is any factorization statistic, then by Theorem 3.2 we have the following evaluation of $E_d(P)$ at $q = 1$,

$$E_d(P)_{q=1} = \sum_{k=0}^{d-1} \langle P, \psi_d^k \rangle = \langle P, \sum_{k=0}^{d-1} \psi_d^k \rangle.$$

Passing to characters in Theorem 4.1 gives

$$\sum_{k=0}^{d-1} \psi_d^k = \chi_{\mathrm{reg}},$$

where χ_{reg} is the character of the regular representation $\mathbb{Q}[S_d]$. If P is a character of an S_d-representation V, then it follows from the general representation theory of finite groups that $\langle P, \chi_{\mathrm{reg}} \rangle = \dim V$. Therefore,

$$E_d(P)_{q=1} = \dim V.$$

Thus if Q is the quadratic excess factorization statistic, then earlier we showed that Q is the character of the $\binom{d}{2}$-dimensional representation $\bigwedge^2 \mathbb{Q}[d]$. Hence

$$E_d(Q)_{q=1} = \binom{d}{2},$$

which was observed in the table of values for $E_d(Q)$. We also showed that the root statistic R was the character of the permutation representation $\mathbb{Q}[d]$, hence

$$E_d(R)_{q=1} = d.$$

Acknowledgements

The author would like to thank Jeff Lagarias for introducing him to splitting measures and for many helpful conversations. Also thanks to Phil Tosteson for valuable references and to Jonathan Gerhard for the observation about specializing expected values at $q = 1$.

References

[1] Weiyan Chen. “Analytic number theory for 0-cycles”. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press. 2017, pp. 1–24.

[2] Thomas Church. “Homological stability for configuration spaces of manifolds”. In: *Inventiones mathematicae* 188.2 (2012), pp. 465–504.

[3] Thomas Church, Jordan Ellenberg, and Benson Farb. “Representation stability in cohomology and asymptotics for families of varieties over finite fields”. In: *Contemporary Mathematics* 620 (2014), pp. 1–54.

[4] Thomas Church and Benson Farb. “Representation theory and homological stability”. In: *Advances in Mathematics* 245 (2013), pp. 250–314.

[5] Patricia Hersh and Victor Reiner. “Representation Stability for Cohomology of Configuration Spaces in”. In: *International Mathematics Research Notices* 2017.5 (2016), pp. 1433–1486.

[6] Trevor Hyde. “Multivariate polynomial factorization statistics and liminal reciprocity”. 2017. In preparation.

[7] Trevor Hyde. “Twisted Grothendieck-Lefschetz formulas and the expected values of polynomial factorization statistics”. 2017. In preparation.

[8] Trevor Hyde and Jeffrey C. Lagarias. “Polynomial factorization measures and Lie algebra cohomology”. 2017. In preparation.

[9] Trevor Hyde and Jeffrey C Lagarias. “Polynomial splitting measures and cohomology of the pure braid group”. In: *Arnold Mathematical Journal* (2017), pp. 1–31.