T. Hyde (2018) "Polynomial Factorization Statistics and Point Configurations in $\mathbb{R}^3$,"
International Mathematics Research Notices, Vol. 00, No. 0, pp. 1–26
Advance Access publication on Month XX, 2018
doi:10.1093/imrn/rny271



# Polynomial Factorization Statistics and Point Configurations in $\mathbb{R}^3$

**Trevor Hyde***

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, USA

**Correspondence to be sent to: e-mail: tghyde@umich.edu*

We use combinatorial methods to relate the expected values of polynomial factorization statistics over $\mathbb{F}_q$ to the cohomology of ordered configurations in $\mathbb{R}^3$ as a representation of the symmetric group. Our method gives a new proof of the twisted Grothendieck–Lefschetz formula for squarefree polynomial factorization statistics of Church, Ellenberg, and Farb.

## 1 Introduction

In this paper we use combinatorial methods from the theory of generating functions to draw a surprising connection between the expected values of arithmetic functions on $\mathbb{F}_q[x]$, combinatorial representation theory, and the cohomology of point configurations in $\mathbb{R}^3$.

**Definition.** Let $\mathrm{Poly}_d(\mathbb{F}_q)$ denote the set of degree $d$ monic polynomials in $\mathbb{F}_q[x]$. The *factorization type* of $f(x) \in \mathrm{Poly}_d(\mathbb{F}_q)$ is the partition of $d$ formed by the degrees of the irreducible factors of $f(x)$ over $\mathbb{F}_q$. A *factorization statistic* $P$ is a function defined on $\mathrm{Poly}_d(\mathbb{F}_q)$ such that $P(f)$ only depends on the factorization type of $f(x)$. Note that $P$ may also be viewed as a function defined on partitions of $d$, or equivalently as a class function of the symmetric group $S_d$. □

Received August 6, 2018; Revised November 2, 2018; Accepted November 5, 2018

**Theorem 1.1.** Let $\psi_d^k$ be the character of the $S_d$-representation $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ where $\mathrm{PConf}_d(\mathbb{R}^3)$ is the ordered configuration space of $d$ distinct points in $\mathbb{R}^3$ (see Section 2.) Then the expected value $E_d(P)$ of a factorization statistic $P$ on $\mathrm{Poly}_d(\mathbb{F}_q)$ is given by

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

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 $\mathbb{Q}$-valued class functions of the symmetric group $S_d$. □

Theorem 1.1 asserts that the expected value of any factorization statistic $P$ on $\mathrm{Poly}_d(\mathbb{F}_q)$ may be expressed as a polynomial in $1/q$ with coefficients determined by the representation theoretic structure of the cohomology of a configuration space in a way that is uniform in $q$. This result provides a bridge between the arithmetic statistics of polynomials over a finite field and the combinatorial topology of the space $\mathrm{PConf}_d(\mathbb{R}^3)$.

As one application of Theorem 1.1 we deduce the following structural description of the total cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$ from a simple probabilistic argument.

**Theorem 1.2.** For each $d \geq 1$ there is an isomorphism of $S_d$-representations

$$\bigoplus_{k=0}^{d-1} H^{2k}\left(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}\right) \cong \mathbb{Q}[S_d], \tag{1.1}$$

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

Theorem 1.2 is known, from other perspectives, to follow from the Poincaré–Birkhoff–Witt theorem [17, p. 56]. We explore consequences of Theorem 1.2 through examples in Section 3.

As a 2nd application of Theorem 1.1 we deduce the asymptotic stability of expected values from the *representation stability* of the family $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ of symmetric group representations.

**Definition.** Let $x_j(f)$ denote the number of degree $j$ irreducible factors of $f \in \mathrm{Poly}_d(\mathbb{F}_q)$. Then a *character polynomial* $P$ is a factorization statistic given by a polynomial in the functions $x_j$ for $j \geq 1$. □

**Theorem 1.3.** Let $P$ be a character polynomial. Then

$$\lim_{d \to \infty} E_d(P) = \sum_{k=0}^{\infty} \frac{\langle P, \psi^k \rangle}{q^k},$$

where the limit is taken $1/q$-adically (or equivalently, coefficientwise in the formal power series ring $\mathbb{Q}[\![1/q]\!]$,) and $\langle P, \psi^k\rangle := \lim_{d\to\infty}\langle P, \psi_d^k\rangle$ is the *stable multiplicity* of $P$ in $\psi_d^k$ (see Section 2.4.) □

The connection between expected values of factorization statistics and the symmetric group representations $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ is made through a probability measure on the symmetric group. Given a partition $\lambda \vdash d$, let $\nu(\lambda)$ denote the probability of a random element of $\mathrm{Poly}_d(\mathbb{F}_q)$ having factorization type $\lambda$. The function $\nu$ is called the *splitting measure*. We prove Theorem 1.4 using a generating function argument in Section 2.

**Theorem 1.4.** Let $\psi_d^k$ be the character of the $S_d$-representation $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$, where $\mathrm{PConf}_d(\mathbb{R}^3)$ is the ordered configuration space of $d$ distinct points in $\mathbb{R}^3$ (see Section 2 for details.) Then for all $d \geq 1$ and partitions $\lambda \vdash d$ we have

$$\nu(\lambda) = \frac{1}{z_\lambda}\sum_{k=0}^{d-1}\frac{\psi_d^k(\lambda)}{q^k},$$

where $z_\lambda := \prod_{j\geq 1} j^{m_j} m_j!$ when $\lambda = (1^{m_1}2^{m_2}\cdots)$, and $\psi_d^k(\lambda)$ is the value of the character $\psi_d^k$ on any element of the symmetric group $S_d$ with cycle type $\lambda$. □

Church *et al.* [8] connect the first moments of factorization statistics on the set $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ of *squarefree* monic degree $d$ polynomials to the symmetric group representations carried by the cohomology of configuration space through their *twisted Grothendieck–Lefschetz formula for* $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$.

**Theorem 1.5** ([8, Prop. 4.1]). Let $\phi_d^k$ be the character of the $S_d$-representation $H^k(\mathrm{PConf}_d(\mathbb{C}), \mathbb{Q})$, where $\mathrm{PConf}_d(\mathbb{C})$ is the ordered configuration space of $d$ distinct points in $\mathbb{C}$. Let $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ denote the set of squarefree monic degree $d$ polynomials in $\mathbb{F}_q[x]$. Then for any factorization statistic $P$,

$$\sum_{f\in\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)} P(f) = q^d \sum_{k=0}^{d-1}\frac{(-1)^k\langle P, \phi_d^k\rangle}{q^k}, \tag{1.2}$$

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

They derive the first moment formula (1.2) from the Grothendieck–Lefschetz trace formula for étale cohomology with "twisted coefficients". Lagarias and the author [16] use Theorem 1.5 to establish a representation theoretic interpretation of the *squarefree splitting measure* $\nu^{\mathrm{sf}}$, where $\nu^{\mathrm{sf}}(\lambda)$ is the probability of a random squarefree polynomial having factorization type $\lambda$.

**Theorem 1.6** ([16, Thm. 1.2]). Let $\chi_d^k$ be the character of the $S_d$-representation $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ (see Section 2.2). Then for all $d \geq 2$ and partitions $\lambda \vdash d$ we have

$$\nu^{\mathrm{sf}}(\lambda) = \frac{1}{z_\lambda} \sum_{k=0}^{d-2} \frac{(-1)^k \chi_d^k(\lambda)}{q^k},$$

where $z_\lambda = \prod_{j\geq 1} j^{m_j} m_j!$ when $\lambda = (1^{m_1} 2^{m_2} \cdots)$, and $\chi_d^k(\lambda)$ is the value of the character $\chi_d^k$ on any element of the symmetric group $S_d$ with cycle type $\lambda$. □

We give a new proof of Theorem 1.6 using the same method as for Theorem 1.4 and derive Theorem 1.5 as a consequence. Our proofs of Theorems 1.1 and 1.5 do not use algebraic geometry or the Grothendieck–Lefschetz trace formula.

The use of generating functions in the study of factorization statistics is not new. Church *et al.* [8] use $L$-functions to compute the stable limits of expected values of squarefree factorization statistics. Fulman [10] uses cycle index series to derive the asymptotic formulas for first moments of squarefree factorization statistics given in [8] without using representation theory or cohomology. Chen [5, 6] further develops these methods in the more general setting of an arbitrary affine or projective variety $V$ defined over $\mathbb{F}_q$. Carlitz [4] uses zeta functions to compute the expected values of specific factorization statistics.

Our main innovation is connecting factorization statistics of polynomials to the cohomology of configurations in $\mathbb{R}^3$ in a way parallel to the connection established by Church *et al.* [8] between factorization statistics of squarefree polynomials and the cohomology of configurations in $\mathbb{C} \cong \mathbb{R}^2$, and providing a unified generating function method to derive both results.

There have been other generalizations of Theorem 1.5 from squarefree polynomials to all polynomials. Gadish [11, Sec. 1.3] and Hast and Matei [13] both study expected values of functions defined on the set of all polynomials; their functions depend on both the degree of the irreducible factors and their multiplicities. We call these *weighted factorization statistics*. Gadish [11, Cor. 1.4] shows that the expected value of a weighted factorization statistic $P$ on $\mathrm{Poly}_d(\mathbb{F}_q)$ matches the expected value of $P$ on $S_d$ viewed as a class function. Stated geometrically, the expected values of weighted factorization statistics on degree $d$ polynomials correspond to the cohomology of $\mathbb{R}^d$ as an $S_d$-representation, while the expected values of our factorization statistics correspond to the cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$ as an $S_d$-representation.

### 1.1 Further questions

Church, Ellenberg, and Farb's étale cohomology approach to Theorem 1.5 illustrates a clear geometric connection between squarefree factorization statistics and the cohomology of ordered configurations in $\mathbb{C}$. To summarize, we start with the map of schemes

$$\mathrm{PConf}_d(\mathbb{A}^1) \longrightarrow \mathrm{Conf}_d(\mathbb{A}^1), \tag{1.3}$$

which sends an ordered configuration of $d$ points to its unordered counterpart. The symmetric group $S_d$ acts freely on $\mathrm{PConf}_d(\mathbb{A}^1)$ by permuting points in the ordered configuration, and the map in (1.3) is the quotient by this action. The $\mathbb{F}_q$-points of $\mathrm{Conf}_d(\mathbb{A}^1)$ are in natural correspondence with squarefree polynomials of degree $d$, and the $\mathbb{C}$-points of $\mathrm{PConf}_d(\mathbb{A}^1)$ give us the manifold $\mathrm{PConf}_d(\mathbb{C})$. The Grothendieck–Lefschetz trace formula connects point counts over finite fields with the étale cohomology of the scheme and general comparison theorems between cohomology theories relate this to the singular cohomology of the manifold $\mathrm{PConf}_d(\mathbb{C})$.

The map (1.3) is unramified, simplifying the application of the Grothendieck–Lefschetz trace formula. The corresponding map of schemes in the case of all polynomials is

$$(\mathbb{A}^1)^d \longrightarrow \mathrm{Sym}_d(\mathbb{A}^1),$$

which is highly ramified. Gadish [11] adapts the étale cohomological perspective to handle ramified covers. This geometrically natural extension leads Gadish to a twisted Grothendieck–Lefschetz formula for weighted factorization statistics [11, Thm. A (1.2)].

Our factorization statistics extend those on $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$ in a way that is combinatorially natural but is difficult to manage from the algebraic geometry perspective. This results in a surprising connection to the cohomology of ordered configurations in $\mathbb{R}^3$ for which we have no geometric explanation.

**Question 1.7.**
Is there a geometric interpretation of the connection between factorization statistics on $\mathrm{Poly}_d(\mathbb{F}_q)$ and the cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$? □

Church, Ellenberg, and Farb deduce their twisted Grothendieck–Lefschetz formula from a more general result relating factorization statistics on quotients of complements of hyperplane arrangements to the étale cohomology of said complements. Note that $\mathrm{PConf}_d(\mathbb{C})$ may be interpreted may be viewed as the complement of the *braid arrangement*, consisting of the hyperplanes $z_i = z_j$ for all $i \neq j$. Given a collection of linear forms $L$ defined over $\mathbb{Z}$ in $d$ variables that is stable under the natural action of $S_d$, let $A_d(L)$ be the complement of the hyperplane arrangement determined by the vanishing sets of the linear forms. Let $B_d(L)$ denote the scheme-theoretic quotient of $A_d(L)$ by the action of $S_d$.

**Theorem 1.8** ([8, Thm. 3.7]). Let $P$ be a factorization statistic. If $\ell$ is a prime not dividing $q$ and $\tau_d^k$ is the $S_d$-character of $H^k_{\text{ét}}(A_d(L), \mathbb{Q}_\ell)$, then

$$\sum_{f \in B(L)_d(\mathbb{F}_q)} P(f) = \sum_{k=0}^{d} (-1)^k \langle P, \tau_d^k \rangle q^{d-k}.$$

Given that our generating function method provides a new proof of the special case Theorem 1.5, we ask the following question.

**Question 1.9.**
Can our methods be adapted to give a new proof of Theorem 1.8? □

The key to answering Question 1.9 is to find explicit product formulas for the cycle index series of the family of representations given by the étale cohomology analogous to those used in our proof of Theorem 2.1. Such formulas may be known, but not to us.

### 1.2 Outline

Section 2 begins with a discussion of splitting measures and a review of the relevant families of representations. Theorems 1.4 and 1.6 are proven together as Theorem 2.1. Section 2.3 introduces factorization statistics and contains our main result on their expected values. The asymptotic stability of expected values for character polynomials is shown as Theorem 2.5. Finally, the cohomology constraint Theorem 1.2 is deduced as Theorem 2.6. Section 3 considers example applications of our results.

## 2 Representation Theoretic Interpretation of Splitting Measures

Let $q$ be a prime power and $d \geq 1$ an integer. Let $\mathrm{Poly}_d(\mathbb{F}_q)$ be the set of monic degree $d$ polynomials in $\mathbb{F}_q[x]$. The subset of squarefree polynomials is denoted $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q) \subseteq \mathrm{Poly}_d(\mathbb{F}_q)$. Every polynomial $f \in \mathrm{Poly}_d(\mathbb{F}_q)$ has a unique factorization into irreducible polynomials over $\mathbb{F}_q$. The degrees of the irreducible factors of $f$ form a partition $[f]$ of the degree $d$ that we call the *factorization type* of $f$. Recall that the number of degree $j$ irreducible polynomials in $\mathbb{F}_q[x]$ is given by the *necklace polynomial*

$$M_j(q) := \frac{1}{j}\sum_{i|j}\mu(i)q^{j/i}.$$

If $\lambda$ is a partition, then $m_j = m_j(\lambda)$ is the number of size $j$ parts of $\lambda$. In other words, $\lambda = (1^{m_1}2^{m_2}3^{m_3}\cdots)$. Given a partition $\lambda \vdash d$ we define $\left(\!\binom{\mathbb{A}^1}{\lambda}\!\right)$ and $\binom{\mathbb{A}^1}{\lambda}$ to be the number of polynomials in $\mathrm{Poly}_d(\mathbb{F}_q)$ and $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$, respectively with factorization type $\lambda$. The $\mathbb{A}^1$ in the notation reflects the correspondence between degree $d$ monic polynomials in $\mathbb{F}_q[x]$ and points in $\mathrm{Sym}_d(\mathbb{A}^1)(\mathbb{F}_q)$. By unique factorization we have the formulas,

$$\left(\!\!\binom{\mathbb{A}^1}{\lambda}\!\!\right) := \prod_{j\geq 1}\left(\!\!\binom{M_j(q)}{m_j}\!\!\right) \qquad \binom{\mathbb{A}^1}{\lambda} := \prod_{j\geq 1}\binom{M_j(q)}{m_j}, \tag{2.1}$$

where

$$\left(\!\!\binom{x}{m}\!\!\right) := \frac{x(x+1)(x+2)\cdots(x+m-1)}{m!} = \binom{x+m-1}{m}.$$

Note that $\left(\!\binom{x}{m}\!\right)$ counts the number of subsets of size $m$ chosen from an $x$ element set with repetition. The product expressions (2.1) show that $\left(\!\binom{\mathbb{A}^1}{\lambda}\!\right)$ and $\binom{\mathbb{A}^1}{\lambda}$ are polynomials in $q$ of degree $d$.

The total number of monic degree $d$ polynomials over $\mathbb{F}_q$ is $|\mathrm{Poly}_d(\mathbb{F}_q)| = q^d$, while the total number of squarefree polynomials is $|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)| = q^d - q^{d-1}$ for $d \geq 2$ (see [18, Prop. 2.3]). We define the *splitting measure* $\nu(\lambda)$ to be the probability of an element $f \in \mathrm{Poly}_d(\mathbb{F}_q)$ having factorization type $\lambda$, and similarly define the *squarefree splitting measure* $\nu^{\mathrm{sf}}(\lambda)$ for $f \in \mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$. More explicitly,

$$\nu(\lambda) := \frac{1}{|\mathrm{Poly}_d(\mathbb{F}_q)|}\left(\!\!\binom{\mathbb{A}^1}{\lambda}\!\!\right) \qquad \nu^{\mathrm{sf}}(\lambda) := \frac{1}{|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)|}\binom{\mathbb{A}^1}{\lambda}.$$

The splitting measure $\nu$ is studied from a statistical point of view in [1].

Both splitting measures are rational functions in $q$ for each partition $\lambda$, and furthermore both are polynomials in $1/q$ (this is clear for $\nu(\lambda)$ and is shown for $\nu^{\mathrm{sf}}(\lambda)$ in [16, Prop. 2.4]). Recall that the partitions $\lambda \vdash d$ parametrize the conjugacy classes of the symmetric group $S_d$. Thus, the splitting measures may be viewed as polynomial-valued class functions on $S_d$. Our 1st result gives an interpretation of the coefficients of the splitting measures in terms of the representation theory of the symmetric group.

Let us review some terminology and notation. If $\chi$ is a character of the symmetric group $S_d$ and $\lambda$ is a partition of $d$, we write $\chi(\lambda)$ for the value of $\chi$ on any element $\sigma \in S_d$ of cycle type $\lambda$. This is well-defined since characters are constant on conjugacy classes. Let $z_\lambda$ be the number of permutations in $S_d$ commuting with an element $\sigma \in S_d$ of cycle type $\lambda$, then

$$z_\lambda = \prod_{j \geq 1} j^{m_j} m_j!\,.$$

The *rank* of a partition $\lambda \vdash d$ is $\mathrm{rk}(\lambda) := \sum_{j \geq 1} m_j - 1 = d - \ell(\lambda)$, where $\ell(\lambda)$ is the number of parts in $\lambda$.

### 2.1 Higher Lie representations

Given a positive integer $j$, let $\zeta_j$ be a faithful one-dimensional complex representation of the cyclic group $C_j$. Viewing $C_j$ as a subgroup of the symmetric group $S_j$ generated by a $j$-cycle, the *jth Lie representation* $\mathrm{Lie}(j)$ is defined as the induced representation

$$\mathrm{Lie}(j) := \mathrm{Ind}_{C_j}^{S_j} \zeta_j.$$

For a partition $\lambda \vdash d$, the *higher Lie representation* $\mathrm{Lie}_\lambda$ is defined as

$$\mathrm{Lie}_\lambda := \mathrm{Ind}_{Z_\lambda}^{S_d} \bigotimes_{j \geq 1} \mathrm{Lie}(j)^{\otimes m_j(\lambda)},$$

where $Z_\lambda$ is the centralizer of a permutation with cycle type $\lambda$. Finally, for $0 \leq k < d$ let $\mathrm{Lie}_d^k$ be the $S_d$-representation

$$\mathrm{Lie}_d^k := \bigoplus_{\mathrm{rk}(\lambda) = k} \mathrm{Lie}_\lambda.$$

### 2.2 Configuration spaces

Given a topological space $X$, let $\mathrm{PConf}_d(X)$ be the space of ordered configurations of $d$ distinct points in $X$,

$$\mathrm{PConf}_d(X) := \big\{(x_1, x_2, \ldots, x_d) \in X^d : x_i \neq x_j \text{ when } i \neq j\big\}.$$

The symmetric group $S_d$ acts freely on $\mathrm{PConf}_d(X)$ by permuting the coordinates. Thus, the singular cohomology $H^k(\mathrm{PConf}_d(X), \mathbb{Q})$ is, by functoriality, an $S_d$-representation for all $k \geq 0$. Sundaram and Welker [19, Thm. 4.4(iii)] show that for $k \geq 0$ and for every odd $n \geq 3$

$$H^{(n-1)k}\big(\mathrm{PConf}_d(\mathbb{R}^n), \mathbb{Q}\big) \cong \mathrm{Lie}_d^k,$$

as $S_d$-representations (see [14, Sec. 2.3] for a discussion of this result in language closer to our presentation). For the sake of concreteness we specialize to the case $n = 3$,

$$H^{2k}\big(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}\big) \cong \mathrm{Lie}_d^k.$$

If $X = \mathbb{C}$, then 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 there is a well-defined $S_d$-action on the quotient $\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times$. Thus, $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$ is an $S_d$-representation for all $k \geq 0$.

We now state our 1st main result.

**Theorem 2.1.** Let $\psi_d^k$ and $\chi_d^k$ be the characters of the $S_d$-representations $\mathrm{Lie}_d^k \cong H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ and $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$, respectively.

1. For $d \geq 1$ and each partition $\lambda \vdash d$,

$$\nu(\lambda) = \frac{1}{z_\lambda} \sum_{k=0}^{d-1} \frac{\psi_d^k(\lambda)}{q^k}.$$

2. For $d \geq 2$ and each partition $\lambda \vdash d$,

$$\nu^{\mathrm{sf}}(\lambda) = \frac{1}{z_\lambda} \sum_{k=0}^{d-2} \frac{(-1)^k \chi_d^k(\lambda)}{q^k}.$$

This representation theoretic interpretation of the squarefree splitting measure was first shown in [16, Thm. 5.1] using the twisted Grothendieck–Lefschetz formula for squarefree factorization statistics of Church *et al.* [8, Prop. 4.1]. We prove Theorem 2.1 using generating functions, leading to a new proof of the twisted Grothendieck–Lefschetz formula for squarefree factorization statistics in Theorem 2.3. The representation theoretic interpretation of the splitting measure $\nu(\lambda)$ appears to be new.

**Proof.** 1. For each integer $j \geq 1$ let $p_j$ be a formal variable. If $\lambda = (1^{m_1}2^{m_2}\cdots)$ is a partition, let $p_\lambda := \prod_{j\geq 1} p_j^{m_j}$. Hersh and Reiner [14, Thm. 2.17] state the following identity of formal power series:

$$\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}\psi_d^k(\lambda)q^{d-k}p_\lambda t^d = \prod_{j\geq 1}\left(\frac{1}{1-p_jt^j}\right)^{M_j(q)}, \tag{2.2}$$

where $M_j(q) = \frac{1}{j}\sum_{i\mid j}\mu(i)q^{j/i}$ is the $j$th necklace polynomial and $\psi_d^k$ is the character of $\mathrm{Lie}_d^k$ (see remarks following the proof for a discussion of the equivalence of (2.2) and [14, Thm. 2.17]). Recall the following version of the binomial theorem for formal power series,

$$\left(\frac{1}{1-t}\right)^m = \sum_{d\geq 0}\left(\!\!\binom{m}{d}\!\!\right)t^d,$$

where $\left(\!\binom{m}{d}\!\right) := \frac{m(m+1)(m+2)\cdots(m+d-1)}{d!}$. Expanding the right-hand side of (2.2) gives

$$\begin{aligned}\prod_{j\geq 1}\left(\frac{1}{1-p_jt^j}\right)^{M_j(q)} &= \prod_{j\geq 1}\sum_{m_j\geq 0}\left(\!\!\binom{M_j(q)}{m_j}\!\!\right)p_j^{m_j}t^{jm_j}\\ &= \sum_{d\geq 0}\sum_{\lambda\vdash d}\left(\prod_{j\geq 1}\left(\!\!\binom{M_j(q)}{m_j(\lambda)}\!\!\right)\right)p_\lambda t^d \\ &= \sum_{d\geq 0}\sum_{\lambda\vdash d}\left(\!\!\binom{\mathbb{A}^1}{\lambda}\!\!\right)p_\lambda t^d.\end{aligned} \tag{2.3}$$

Substitute $t = 1/q$ in (2.2) and (2.3) to arrive at

$$\begin{aligned}\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}\frac{\psi_d^k(\lambda)}{q^k}p_\lambda &= \prod_{j\geq 1}\left(\frac{1}{1-p_j/q^j}\right)^{M_j(q)}\\ &= \sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{q^d}\left(\!\!\binom{\mathbb{A}^1}{\lambda}\!\!\right)p_\lambda\\ &= \sum_{d\geq 0}\sum_{\lambda\vdash d}\nu(\lambda)p_\lambda.\end{aligned}$$

Comparing coefficients of $p_\lambda$ we conclude that

$$\nu(\lambda) = \frac{1}{z_\lambda}\sum_{k=0}^{d-1}\frac{\psi_d^k(\lambda)}{q^k}.$$

2. The derivation of the formula for $\nu^{\mathrm{sf}}(\lambda)$ starts with another formal power series identity from [14, Thm. 2.17]. Let $\phi_d^k$ be the character of the $S_d$-representation $H^k(\mathrm{PConf}_d(\mathbb{C}),\mathbb{Q})$. Then

$$\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}\phi_d^k(\lambda)q^{d-k}p_\lambda t^d=\prod_{j\geq 1}\left(1+(-1)^j p_j t^j\right)^{M_j(-q)}.$$

The substitution $t\mapsto -t$ and $q\mapsto -q$ simplifies this to

$$\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}(-1)^k\phi_d^k(\lambda)q^{d-k}p_\lambda t^d=\prod_{j\geq 1}\left(1+p_j t^j\right)^{M_j(q)}. \tag{2.4}$$

By the binomial theorem, the right-hand side of (2.4) expands as

$$\begin{aligned}\prod_{j\geq 1}(1+p_jt^j)^{M_j(q)}&=\prod_{j\geq 1}\sum_{m_j\geq 0}\binom{M_j(q)}{m_j}p_j^{m_j}t^{jm_j}\\&=\sum_{d\geq 0}\sum_{\lambda\vdash d}\left(\prod_{j\geq 1}\binom{M_j(q)}{m_j(\lambda)}\right)p_\lambda t^d\\&=\sum_{d\geq 0}\sum_{\lambda\vdash d}\binom{\mathbb{A}^1}{\lambda}p_\lambda t^d.\end{aligned} \tag{2.5}$$

Substituting $t=1/q$ in (2.4) and (2.5) gives

$$\begin{aligned}\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}\frac{\phi_d^k(\lambda)}{q^k}p_\lambda&=\prod_{j\geq 1}\left(1+p_j/q^j\right)^{M_j(q)}\\&=\sum_{d\geq 0}\sum_{\lambda\vdash d}\frac{1}{q^d}\binom{\mathbb{A}^1}{\lambda}p_\lambda.\end{aligned} \tag{2.6}$$

Let $\chi_d^k$ be the character of the $S_d$-representation $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times,\mathbb{Q})$. Hyde and Lagarias [16, Prop. 4.2 and Thm. 4.3] constructed an isomorphism of $S_d$-representations

$$H^k(\mathrm{PConf}_d(\mathbb{C}),\mathbb{Q})\cong H^k\left(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times,\mathbb{Q}\right)\oplus H^{k-1}\left(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times,\mathbb{Q}\right),$$

from which it follows that $\phi_d^k = \chi_d^k + \chi_d^{k-1}$.

Note that $H^{-1}(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q}) = H^{d-1}(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q}) = 0$. Therefore,

$$\begin{aligned}\frac{1}{1-\frac{1}{q}}\sum_{k=0}^{d-1}\frac{(-1)^k\phi_d^k(\lambda)}{q^k} &= \frac{1}{1-\frac{1}{q}}\sum_{k=0}^{d-1}\frac{(-1)^k\left(\chi_d^k(\lambda)+\chi_d^{k-1}(\lambda)\right)}{q^k}\\ &= \frac{1}{1-\frac{1}{q}}\sum_{k=0}^{d-2}\frac{(-1)^k\chi_d^k(\lambda)}{q^k}+\frac{(-1)^{k+1}\chi_d^k(\lambda)}{q^{k+1}}\\ &= \sum_{k=0}^{d-2}\frac{(-1)^k\chi_d^k(\lambda)}{q^k}.\end{aligned} \tag{2.7}$$

Multiplying (2.6) by $\frac{1}{1-\frac{1}{q}}$ for $d \geq 2$ leads to

$$\begin{aligned}\sum_{d\geq 2}\sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-2}\frac{(-1)^k\chi_d^k(\lambda)}{q^k}p_\lambda &= \sum_{d\geq 2}\sum_{\lambda\vdash d}\frac{1}{q^d-q^{d-1}}\binom{\mathbb{A}^1}{\lambda}p_\lambda\\ &= \sum_{d\geq 2}\sum_{\lambda\vdash d}\nu^{\mathrm{sf}}(\lambda)p_\lambda.\end{aligned}$$

Finally, comparing coefficients of $p_\lambda$ we conclude that for $d \geq 2$

$$\nu^{\mathrm{sf}}(\lambda) = \frac{1}{z_\lambda}\sum_{k=0}^{d-2}\frac{(-1)^k\chi_d^k(\lambda)}{q^k}.$$

■

The generating functions used in the proof of Theorem 2.1 are stated in terms of symmetric functions in [14]. To convert between their notation and ours one can interpret their power symmetric functions $p_j$ as formal variables, and their Frobenius characteristic $\mathrm{ch}(V)$ of an $S_d$-representation $V$ with character $\chi_V$ by its *cycle indicator*

$$\mathrm{ch}(V) \longrightarrow \sum_{\lambda\vdash d}\frac{\chi_V(\lambda)}{z_\lambda}p_\lambda.$$

Hersh and Reiner cite several sources for the origin of these generating functions. A derivation of the identity for the higher Lie characters may be found in [12, Thm. 3.7], although the characters are not called by this name there. The generating function for the cohomology of configurations in $\mathbb{C}$ is derived in [2, Cor. 4.4] with notation similar to ours but stated in a way that does not explicitly connect it with configuration space. Both product formulas result from a plethystic decomposition of the respective families of representations.

### 2.3 Factorization statistics and the cohomology of configuration space

A *factorization statistic* $P$ is a function defined on $\mathrm{Poly}_d(\mathbb{F}_q)$ such that $P(f)$ only depends on the factorization type of $f \in \mathrm{Poly}_d(\mathbb{F}_q)$. Equivalently, $P$ may be viewed as a function defined on the set of partitions of $d$ or as a class function of the symmetric group $S_d$. Any class function $P$ may be interpreted as a factorization statistic.

**Example 2.1.** We illustrate 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 $(1^3\,2^1)$ and the factorization type of $h(x)$ is $(1^2\,3^1)$. Note that the factorization type does not detect the multiplicity of factors so that $x^2$ and $x(x+1)$ both have the same factorization type $(1^2)$.

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, $R(g) = 3$ and $R(h) = 2$.
3. For $k \geq 1$, let $x_k(f)$ be the number of degree $k$ irreducible factors of $f \in \mathrm{Poly}_d(\mathbb{F}_q)$, then $x_k$ is a factorization statistic. As a function on partitions $x_k(\lambda) = m_k(\lambda)$ is the number of parts of $\lambda$ of size $k$. Note that $R = x_1$. The ring $\mathbb{Q}[x_1, x_2, \ldots]$ generated by the functions $x_k$ for $k \geq 1$ is called the ring of *character polynomials*. We return to character polynomials in Section 2.4 when discussing asymptotic stability.
4. Say a polynomial $f(x)$ has *even type* if the factorization type of $f(x)$ is an even partition. In other words, suppose $\lambda = (1^{m_1} 2^{m_2} 3^{m_3} \cdots)$ 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 indicator function $ET$ defined by

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

is a factorization statistic. We compute $ET(g) = 0$ and $ET(h) = 1$.

Let $E_d(P)$ denote the expected value of a factorization statistic $P$ on $\mathrm{Poly}_d(\mathbb{F}_q)$ and let $E_d^{\mathrm{sf}}(P)$ denote the expected value of $P$ on $\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)$. More precisely,

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

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

**Example 2.2** (Quadratic excess). This example is inspired by [8, p. 6]. Define the *quadratic excess* $Q(f)$ of a polynomial $f(x) \in \mathbb{F}_q[x]$ to be

$$\begin{aligned} Q(f) =&\#\{\text{reducible quadratic factors of } f(x)\} \\ &- \#\{\text{irreducible quadratic factors of } f(x)\}, \end{aligned}$$

where both counts are considered with multiplicity. Note that $Q(f)$ depends only on the number of linear and irreducible quadratic factors of $f(x)$. For instance, if $g(x) = x^2(x+1)(x^2+1)^4 \in \mathbb{F}_3[x]$, then $g(x)$ has three linear factors and four irreducible quadratic factors, hence

$$Q(g) = \binom{3}{2} - \binom{4}{1} = -1.$$

The table below gives the expected value $E_d(Q)$ for small values of $d$.

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

We note a few remarkable features of these expected values. For each $d$, $E_d(Q)$ 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. Evaluating the polynomial $E_d(Q)$ at $q = 1$ gives the binomial coefficient $\binom{d}{2}$. 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 $d \to \infty$.

All of these observations are verified and explained by Theorem 2.2. We return to this example in Section 3.1.

Our 2nd main result gives an explicit expression for the expected value $E_d(P)$ of a factorization statistic in terms of the ordered configuration space of $d$ distinct points in $\mathbb{R}^3$.

If $P$ and $Q$ are $\mathbb{Q}$-valued class functions on $S_d$, let $\langle P, Q\rangle$ denote their standard $S_d$-invariant inner product

$$\langle P, Q\rangle := \frac{1}{d!}\sum_{\sigma\in S_d} P(\sigma)Q(\sigma) = \sum_{\lambda\vdash d}\frac{P(\lambda)Q(\lambda)}{z_\lambda}.$$

**Theorem 2.2.** Suppose $P$ is a factorization statistic and $d \geq 1$. If $\psi_d^k$ is the character of the $S_d$-representation $\mathrm{Lie}_d^k \cong H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$, then

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

**Proof.** Since factorization statistics depend only on the factorization type of a polynomial, the expected value $E_d(P)$ may be written in terms of the splitting measure as

$$E_d(P) = \frac{1}{|\mathrm{Poly}_d(\mathbb{F}_q)|}\sum_{f\in\mathrm{Poly}_d(\mathbb{F}_q)} P(f) = \sum_{\lambda\vdash d} P(\lambda)\nu(\lambda).$$

Then Theorem 2.1 implies

$$\begin{aligned}
E_d(P) &= \sum_{\lambda\vdash d} P(\lambda)\nu(\lambda)\\
&= \sum_{\lambda\vdash d}\frac{1}{z_\lambda}\sum_{k=0}^{d-1}\frac{P(\lambda)\psi_d^k(\lambda)}{q^k}\\
&= \sum_{k=0}^{d-1}\frac{1}{q^k}\left(\sum_{\lambda\vdash d}\frac{P(\lambda)\psi_d^k(\lambda)}{z_\lambda}\right)\\
&= \sum_{k=0}^{d-1}\frac{\langle P, \psi_d^k\rangle}{q^k}.
\end{aligned}$$

■

Church, Ellenberg, and Farb relate the first moments of factorization statistics on squarefree polynomials to the ordered configuration space of $d$ distinct points in $\mathbb{C}$ in [8]. Let $\phi_d^k$ be the character of $H^k(\mathrm{PConf}_d(\mathbb{C}), \mathbb{Q})$ as a representation of $S_d$. In [8, Prop. 4.1], Church *et al.* show that

$$\sum_{f \in \mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)} P(f) = \sum_{k=0}^{d-1} (-1)^k \langle P, \phi_d^k \rangle q^{d-k}. \tag{2.8}$$

Dividing by $|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)| = q^d - q^{d-1}$ gives the expected value, but also changes the coefficients on the right-hand side. The calculation (2.7) in the proof of Theorem 2.1 shows that the identity (2.8) is equivalent to Theorem 2.3 below.

We give a new proof of [8, Prop. 4.1] using Theorem 2.1.

**Theorem 2.3.** Suppose $P$ is a factorization statistic and $d \geq 2$. If $\chi_d^k$ is the character of the $S_d$-representation $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q})$, then

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

**Proof.** The proof is parallel to that of Theorem 2.2. First note that

$$E_d^{\mathrm{sf}}(P) = \frac{1}{|\mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)|} \sum_{f \in \mathrm{Poly}_d^{\mathrm{sf}}(\mathbb{F}_q)} P(f) = \sum_{\lambda \vdash d} P(\lambda) \nu^{\mathrm{sf}}(\lambda),$$

and then use Theorem 2.1 to conclude

$$\begin{aligned} E_d^{\mathrm{sf}}(P) &= \sum_{\lambda \vdash d} P(\lambda) \nu^{\mathrm{sf}}(\lambda) \\ &= \sum_{\lambda \vdash d} \frac{1}{z_\lambda} \sum_{k=0}^{d-2} \frac{(-1)^k P(\lambda) \chi_d^k(\lambda)}{q^k} \\ &= \sum_{k=0}^{d-2} \frac{(-1)^k}{q^k} \left( \sum_{\lambda \vdash d} \frac{P(\lambda) \chi_d^k(\lambda)}{z_\lambda} \right) \\ &= \sum_{k=0}^{d-2} \frac{(-1)^k \langle P, \chi_d^k \rangle}{q^k}. \end{aligned}$$

■

The étale cohomological approach to Theorem 2.3 taken in [8] connects squarefree polynomials over $\mathbb{F}_q$ with the configuration space of points on the affine line. The geometric perspective seems to break down in the case of Theorem 2.2. There is no apparent correspondence between configurations of distinct points in $\mathbb{R}^3$ and monic polynomials over $\mathbb{F}_q$. We would be interested to know if there is a geometric explanation for the relationship between the representations $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ and the expected value of factorization statistics on $\mathrm{Poly}_d(\mathbb{F}_q)$.

### 2.4 Asymptotic stability

Church [7, Thm. 1] showed that for all $k \geq 0$ and $n \geq 2$ the families of symmetric group representations $H^k(\mathrm{PConf}_d(\mathbb{R}^n), \mathbb{Q})$ are *representation stable*. We do not require the details of representation stability (the interested reader should consult [9]), only the following fact [8, Sec. 3.4] that we take as a black box: if $P$ is a factorization statistic given by a character polynomial (see Example 2.1 (3)) and $A_d$ is a sequence of $S_d$-representations with characters $\alpha_d$ that exhibit "representation stability", then the sequence of inner products $\langle P, \alpha_d\rangle$ is eventually constant. In that case we write $\langle P, \alpha\rangle$ for the limit of $\langle P, \alpha_d\rangle$ as $d \to \infty$.

Church, Ellenberg, and Farb use the representation stability of $H^k(\mathrm{PConf}_d(\mathbb{C}), \mathbb{Q})$ to prove Theorem 2.4.

**Theorem 2.4** ([8, Thm. 1]). Let $P$ be a factorization statistic given by a character polynomial and write $\langle P, \phi^k\rangle$ for the limit of $\langle P, \phi^k_d\rangle$ as $d \to \infty$. Then

$$\lim_{d\to\infty} \frac{1}{q^d} \sum_{f\in \mathrm{Poly}^{\mathrm{sf}}_d(\mathbb{F}_q)} P(f) = \sum_{k=0}^{\infty} \frac{(-1)^k \langle P, \phi^k\rangle}{q^k}.$$

Church's theorem implies that for each $k$, $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ is representation stable. Hyde and Lagarias showed that $H^k(\mathrm{PConf}_d(\mathbb{C})/\mathbb{C}^\times, \mathbb{Q}) \cong \beta_{[k]}(\Pi_d)$ as $S_d$-representations where $\beta_{[k]}(\Pi_d)$ are the *rank-selected homology of the partition lattice*. Hersh and Reiner [14, Thm. 1.8] showed that $\beta_{[k]}(\Pi_d)$ is representation stable. Therefore, we deduce the asymptotic stability of expected values from Theorems 2.2 and 2.3.

**Theorem 2.5** (Asymptotic stability of expected values). Let $P$ be a factorization statistic given by a character polynomial (see Section 2.1(3)). Then

$$\lim_{d\to\infty} E_d(P) = \sum_{k=0}^{\infty} \frac{\langle P, \psi^k\rangle}{q^k} \qquad \lim_{d\to\infty} E^{\mathrm{sf}}_d(P) = \sum_{k=0}^{\infty} \frac{(-1)^k\langle P, \chi^k\rangle}{q^k},$$

where the limits are taken $1/q$-adically (or equivalently coefficientwise in $\mathbb{Q}[\![1/q]\!]$.)

### 2.5 Constraint on $E_d(P)$ coefficients

Theorem 2.6 below identifies the total cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$ with the regular representation $\mathbb{Q}[S_d]$.

**Theorem 2.6.** For each $d \geq 1$ there is an isomorphism of $S_d$-representations

$$\bigoplus_{k=0}^{d-1} H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}) \cong \mathbb{Q}[S_d], \tag{2.9}$$

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

**Proof.** Let $\rho$ be the character of $\bigoplus_{k=0}^{d-1} H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$. Then

$$\rho = \sum_{k=0}^{d-1} \psi_d^k,$$

where $\psi_d^k$ is the character of $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$. It suffices to show that $\rho$ is equal to the character of the regular representation, that is

$$\rho(\lambda) = \begin{cases} d! & \lambda = [1^d] \\ 0 & \text{otherwise.} \end{cases}$$

By Theorem 2.1 we have

$$\nu(\lambda) = \frac{1}{z_\lambda} \sum_{k=0}^{d-1} \frac{\psi_d^k(\lambda)}{q^k},$$

where $\nu$ is the splitting measure defined by

$$\nu(\lambda) = \frac{1}{q^d} \prod_{j \geq 1} \left(\!\!\binom{M_j(q)}{m_j}\!\!\right).$$

Let $\nu_1$ denote the splitting measure evaluated at $q = 1$. Then $\nu_1(\lambda) = \frac{\rho(\lambda)}{z_\lambda}$. On the other hand, $M_j(1) = 0$ for $j > 1$ and $M_1(1) = 1$ so

$$\nu_1(\lambda) = \prod_{j \geq 1} \left(\!\!\binom{M_j(1)}{m_j}\!\!\right) = \begin{cases} 1 & \lambda = [1^d] \\ 0 & \text{otherwise.} \end{cases}$$

Since $z_{[1^d]} = d!$ the result follows. ■

The following corollary will be used in Section 3 to explain common phenomena that arise in expected value computations for factorization statistics.

**Corollary 2.7.** Suppose $P$ is a factorization statistic defined on $\mathrm{Poly}_d(\mathbb{F}_q)$, which viewed as a class function of $S_d$, is the character of an $S_d$-representation $V$. Let $E_d(P)$ be the expected value of $P$ on $\mathrm{Poly}_d(\mathbb{F}_q)$.

1. $E_d(P)$ is a polynomial in $1/q$ of degree at most $d-1$ with nonnegative integer coefficients.
2. The evaluation of $E_d(P)$ at $q=1$ is $E_d(P)_{q=1} = \dim V$.

**Proof.** 1. Recall that the inner product $\langle \chi, \psi \rangle$ of characters is the dimension of the vector space of maps between the corresponding representations, hence is a non-negative integer. Thus, if $P$ is an $S_d$-character then Theorem 2.2 implies that

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

has nonnegative coefficients.

2. The inner product of class functions is bilinear. Therefore, by Theorem 2.6

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

It follows from the general representation theory of finite groups that $\langle P, \chi_{\mathrm{reg}} \rangle = \dim V$. Therefore,

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

∎

## 3 Examples

Theorems 2.2 and 2.3 form a bridge connecting polynomial factorization statistics on the one hand and representations of the symmetric group and cohomology of configuration spaces on the other. Translating information back and forth across this bridge leads to an interesting interplay among these structures. In this section we first revisit the example of quadratic excess $Q$ to see how our results explain the properties of $E_d(Q)$ observed in the introduction. We finish with some results on expected values and the structure of $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ using the constraint provided by Theorem 2.6.

### 3.1 Quadratic excess

Recall the quadratic excess factorization statistic $Q$ from Example 2.2. $Q(f)$ is defined as the difference between the number of reducible versus irreducible quadratic factors of $f$. Rephrasing this in terms of partitions, if $x_k(\lambda)$ is the number of parts of $\lambda$ of size $k$, then

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

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 2nd exterior power $\bigwedge^2 \mathbb{Q}[d]$. This representation has dimension $\binom{d}{2}$ with a natural basis given by

$$\{e_i \wedge e_j : i < j\}.$$

If $\sigma \in S_d$ is a permutation, then the trace of $\sigma$ on $\bigwedge^2 \mathbb{Q}[d]$ is

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

Thus, $Q$ viewed as a class function of $S_d$, is the character of $\bigwedge^2 \mathbb{Q}[d]$. It follows from Corollary 2.7 that coefficients of $E_d(Q)$ are nonnegative integers summing to $\binom{d}{2} = \dim \bigwedge^2 \mathbb{Q}[d]$. The coefficientwise convergence of $E_d(Q)$ follows from Theorem 2.5. The $1/q$-adic limit of $E_d(Q)$ as $d \to \infty$ is a rational function of $q$, which explains the simple pattern emerging in the coefficients of $E_d(q)$. In particular, using [6, Cor. 10] we compute,

$$\begin{aligned} \lim_{d\to\infty} E_d(Q) &= \frac{1}{2}\left(1 + \frac{1}{q}\right)\left(\frac{1}{1-\frac{1}{q}}\right)^2 - \frac{1}{2}\left(1 - \frac{1}{q}\right)\left(\frac{1}{1-\frac{1}{q^2}}\right) \\ &= \tfrac{2}{q} + \tfrac{2}{q^2} + \tfrac{4}{q^3} + \tfrac{4}{q^4} + \tfrac{6}{q^5} + \tfrac{6}{q^6} + \tfrac{8}{q^7} + \tfrac{8}{q^8} + \tfrac{10}{q^9} + \cdots. \end{aligned}$$

### 3.2 Identifying irreducible components

Theorem 2.6 gives a constraint on the cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$,

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

where $\mathbb{Q}[S_d]$ is the regular representation of the symmetric group. The regular representation of $S_d$ 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 2.6 tells us that the irreducible components $\mathcal{S}_\lambda$ of $\mathbb{Q}[S_d]$ are distributed among the various degrees of cohomology on the left-hand side of (2.9). Theorem 2.2 implies that the filtration of the regular representation given by Theorem 2.6 completely determines and is determined by the expected values of factorization statistics on $\mathrm{Poly}_d(\mathbb{F}_q)$. We use Theorem 2.6 to identify the degrees of some of the irreducible $S_d$-representations in the cohomology of $\mathrm{PConf}_d(\mathbb{R}^3)$.

### 3.3 Trivial representation

Let $\mathbf{1} = \mathcal{S}_{[d]}$ be the one-dimensional *trivial representation* of $S_d$. The character of the trivial representation is constant equal to 1. Interpreting the trivial character as a factorization statistic we have $E_d(1) = 1$ and Theorem 2.2 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 $\langle 1, \psi_d^0 \rangle = 1$ and $\langle 1, \psi_d^k \rangle = 0$ for $k > 0$. Hence, $\mathbf{1}$ 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 $H^0(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ is one-dimensional. Thus,

$$H^0\big(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q}\big) \cong \mathbf{1}, \tag{3.1}$$

and $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ has no trivial component for $k > 0$.

Recall that the characters $\chi_\lambda$ of the irreducible representations $\mathcal{S}_\lambda$ of $S_d$ form a $\mathbb{Q}$-basis for the vector space of all class functions. If $P$ is a factorization statistic, then there are $a_\lambda(P) \in \mathbb{Q}$ such that

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

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

**Corollary 3.1.** If $P$ is any factorization statistic and $a_1(P)$ is the coefficient of the trivial character in the expression of $P$ as a linear combination of irreducible $S_d$-characters, then

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

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

### 3.4 Sign representation

Let $\mathbf{Sgn}_d := \mathcal{S}_{[1^d]}$ be the one-dimensional *sign representation*. The character of $\mathbf{Sgn}_d$ is $\mathrm{sgn}_d(\lambda) = (-1)^{d-\ell(\lambda)}$, or equivalently $\mathrm{sgn}_d([j]) = (-1)^{j-1}$ for a partition $[j]$ with one part of size $j$ and then $\mathrm{sgn}_d$ extends multiplicatively to partitions with more than one part. Viewing $\mathrm{sgn}_d$ as a factorization statistic Theorem 2.2 implies

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

On the other hand, Corollary 2.7 tells us that $\langle \mathrm{sgn}_d, \psi_d^k \rangle = 1$ for exactly one $k$ and is 0 otherwise—which value of $k$ is it?

**Theorem 3.2.** For each $d \geq 1$,

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

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

We prove Theorem 3.2 in [15] using *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.

Recall that the *Liouville function* $\lambda(f)$ is defined to be $-1$ if $f$ is irreducible and extended multiplicatively. Note that $\lambda(f) = (-1)^d \mathrm{sgn}_d(f)$. Carlitz [4, Sec. 3] computed the expected value of the Liouville function on $\mathrm{Poly}_d(\mathbb{F}_q)$ using zeta functions, and Theorem 3.2 may also be deduced from his result. See the announcement [3, p. 121] for a clear statement of his result.

Theorem 3.2 has a surprising consequence. Recall that the *even type* factorization statistic $ET$ is defined by $ET(f) = 1$ when the factorization type of $f$ is an even partition and $ET(f) = 0$ 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. 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, matching our expectation, but there is a slight 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.

### 3.5 Standard representation

Let $\mathbb{Q}[d]$ be the permutation representation of $S_d$. 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 $(d-1)$-dimensional *standard representation* of $S_d$. Let $R$ be the character of $\mathbb{Q}[d]$. If $\sigma \in S_d$, then $R(\sigma)$ is the number of fixed points of $\sigma$ acting on the set $\{1, 2, \ldots, d\}$; hence, $R(\lambda) = x_1(\lambda)$ is the number of parts of $\lambda$ of size one. Viewed as a factorization statistic, $R(f)$ counts the number of $\mathbb{F}_q$-roots of $f$ with multiplicity.

**Theorem 3.3.** Let $R(f)$ be the number of $\mathbb{F}_q$-roots with multiplicity of $f \in \mathrm{Poly}_d(\mathbb{F}_q)$. Then the expected value $E_d(R)$ of $R$ on $\mathrm{Poly}_d(\mathbb{F}_q)$ is

$$E_d(R) = \frac{1 - \frac{1}{q^d}}{1 - \frac{1}{q}} = 1 + \frac{1}{q} + \frac{1}{q^2} + \frac{1}{q^3} + \ldots + \frac{1}{q^{d-1}}. \tag{3.2}$$

It follows that the multiplicity of $\mathbf{Std}$ in $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ is 1 for $0 < k < d$. □

**Proof.** First note that

$$E_d(R) = \frac{1}{q^d} \sum_{f \in \mathrm{Poly}_d(\mathbb{F}_q)} R(f) = \sum_{\lambda \vdash d} x_1(\lambda)\nu(\lambda),$$

where $\nu$ is the splitting measure. In the course of proving Theorem 2.1 we derived the following formal power series identity,

$$\sum_{d \geq 0} \sum_{\lambda \vdash d} \nu(\lambda) p_\lambda = \prod_{j \geq 1} \left( \frac{1}{1 - p_j/q^j} \right)^{M_j(q)}. \tag{3.3}$$

Consider the effect of the operator $p_1 \frac{\partial}{\partial p_1}$ on (3.3). On the left-hand side we get

$$p_1 \frac{\partial}{\partial p_1} \sum_{d\geq 0} \sum_{\lambda \vdash d} \nu(\lambda) p_\lambda = \sum_{d\geq 1} \sum_{\lambda \vdash d} x_1(\lambda) \nu(\lambda) p_\lambda.$$

On the right-hand side we have

$$p_1 \frac{\partial}{\partial p_1} \prod_{j\geq 1} \left( \frac{1}{1 - p_j/q^j} \right)^{M_j(q)} = \frac{M_1(q) p_1}{q(1 - p_1/q)} \prod_{j\geq 1} \left( \frac{1}{1 - p_j/q^j} \right)^{M_j(q)}.$$

Now substitute $p_j \mapsto t^j$ for all $j$ to arrive at

$$\sum_{d\geq 1} \sum_{\lambda \vdash d} x_1(\lambda) \nu(\lambda) t^d = \sum_{d\geq 1} E_d(R) t^d,$$

on the left and

$$\begin{aligned} \frac{M_1(q) t}{q(1 - t/q)} \prod_{j\geq 1} \left( \frac{1}{1 - t^j/q^j} \right)^{M_j(q)} &= \frac{t}{1 - t/q} \prod_{j\geq 1} \left( \frac{1}{1 - (t/q)^j} \right)^{M_j(q)} \\ &= \frac{t}{1 - t/q} \cdot \frac{1}{1 - t} \end{aligned}$$

on the right, where the last equality is a consequence of the *cyclotomic identity*:

$$\frac{1}{1 - qt} = \prod_{j\geq 1} \left( \frac{1}{1 - t^j} \right)^{M_j(q)}.$$

Together these give

$$\sum_{d\geq 1} E_d(R) t^d = \frac{t}{1 - t/q} \cdot \frac{1}{1 - t}. \tag{3.4}$$

Expanding the right-hand side of (3.4) we find

$$\frac{t}{1 - t/q} \cdot \frac{1}{1 - t} = \frac{1}{1 - t} \sum_{d\geq 1} \frac{1}{q^{d-1}} t^d = \sum_{d\geq 1} \left( \frac{1 - \frac{1}{q^d}}{1 - \frac{1}{q}} \right) t^d.$$

Comparing coefficients of $t^d$ we conclude that

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

The assertions about the multiplicity of **Std** in $H^{2k}(\mathrm{PConf}_d(\mathbb{R}^3), \mathbb{Q})$ follow from Theorem 2.2 and (3.1). ■


## Acknowledgments

We are gracious to Jeff Lagarias for asking the question whether there may be interesting structure in the splitting measure of $\mathrm{Poly}_d(\mathbb{F}_q)$. We also thank Weiyan Chen, Nir Gadish, Ofir Gorodetsky, Will Sawin, Phil Tosteson, and Michael Zieve for helpful conversations, references, and feedbacks.


## References


[1] Arratia, R., A. D. Barbour, and S. Tavaré. "On random polynomials over finite fields". *Math. Proc. Camb. Philos. Soc.* 114, no. 2 (1993): 347–68.

[2] Calderbank, A. R., P. Hanlon, and R. W. Robinson. "Partitions into even and odd block size and some unusual characters of the symmetric groups." *Proc. Lond. Math. Soc.* 3, no. 2 (1986): 288–320.

[3] Carlitz, L. "The arithmetic of polynomials in a Galois field, *Proc. Natl. Acad. Sci. USA*, 17, no. 2 (1931): 120–22.

[4] Carlitz, L. "The arithmetic of polynomials in a Galois field." *Amer. J. Math.* 54, no. 1 (1932): 39–50.

[5] Chen, W. "Analytic number theory for 0-cycles." *Math. Proc. Camb. Philos. Soc.* (2017): 1–24.

[6] Chen, W. *Twisted cohomology of configuration spaces and spaces of maximal tori via point-counting*. arXiv:1603.03931 (2016).

[7] Church, T. "Homological stability for configuration spaces of manifolds". *Invent. Math.* 188 (2012): 465–504.

[8] Church, T., J. Ellenberg, and B. Farb. "Representation stability in cohomology and asymptotics for families of varieties over finite fields". *Contemp. Math.* 620 (2014): 1–54.

[9] Church, T. and Farb, B. "Representation theory and homological stability". *Adv. Math.* 245, (2013): 250–314.

[10] Fulman, J. "A generating function approach to counting theorems for square-free polynomials and maximal tori". *Ann. Comb.* 20, (2016): 587–99.

[11] Gadish, N. "A trace formula for the distribution of rational $G$-orbits in ramified covers, adapted to representation stability". *New York J. Math.* 23, (2017): 987–1011.

[12] Hanlon, P. "The action of $S_n$ on the components of the Hodge decomposition of Hochschild homology". *Michigan Math. J.* 37 (1990): 105–24.

[13] Hast, D. and V. Matei. "Higher moments of arithmetic functions in short intervals: a geometric perspective". *Int. Math. Res. Not.* (2018).

[14] Hersh, P. and V. Reiner. "Representation stability for cohomology of configuration spaces in $\mathbb{R}^d$". *Int. Math. Res. Not.* 2017, no. 5 (2017): 1433–86.

[15] Hyde, T. *"Liminal factorization statistics and reciprocity", to appear in Alg. Comb.*, (2018).

[16] Hyde, T. and J. C. Lagarias. "Polynomial splitting measures and cohomology of the pure braid group". *Arnold Math. J.* 3, no. 2 (2017): 219–49.

[17] Reiner, V., F. Saliola, and V. Welker. "Spectra of symmetrized shuffling operators". *Amer. Math. Soc.* 228, no. 1072 (2014).

[18] Rosen, M. *Number Theory in Function Fields*. *Graduate Texts in Mathematics* 120. New York: Springer Science & Business Media, 2013.

[19] Sundaram, S. and V. Welker. "Group actions on arrangements of linear subspaces and applications to configuration spaces". *Trans. Amer Math. Soc.* 349, no. 4 (1997): 1389–420.