DISTRIBUTION OF THE NUMBER OF ZEROS OF POLYNOMIALS OVER A FINITE FIELD

RITIK JAIN, HAN-BOM MOON, AND PETER WU

ABSTRACT. We study the probability distribution of the number of zeros of multivariable polynomials with bounded degree over a finite field. We find the probability generating function for each set of bounded degree polynomials. In particular, in the single variable case, we show that as the degree of the polynomials and the order of the field simultaneously approach infinity, the distribution converges to a Poisson distribution.

1. INTRODUCTION

For some prime number p, take a random polynomial $f \in \mathbb{Z}/p\mathbb{Z}[x]$ with positive degree d. What is the average number of distinct zeros of f in $\mathbb{Z}/p\mathbb{Z}$? While it is clearly somewhere between 0 and p, the exact answer is not quite obvious. Interestingly, the average is *always* one, regardless of the degree d and the prime p. The authors believe this fact is well known in the algebraic combinatorics community.

In this paper, extending the above result, we investigate the distribution of the number of distinct zeros of a random polynomial over any finite field. Let $\mathbb{F}_q$ be the finite field of order $q = p^r$, where p is a prime number. Fix a nonnegative integer d, and consider the set $\mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$ of polynomials with n variables, $\mathbb{F}_q$-coefficients, and degree at most d with respect to each variable. This is a $(d+1)^n$-dimensional $\mathbb{F}_q$-vector space. By selecting coefficients uniformly, we choose a random polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$. Let $X_{n,d} : \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)} \to \mathbb{N}$ be the random variable of the number of distinct zeros of f. The main result of this paper is the computation of its probability generating function.

Theorem 1.1 (Theorem 4.5). *Let $X_{n,d} : \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)} \to \mathbb{N}$ be the random variable whose value is the number of distinct zeros of a random polynomial over $\mathbb{F}_q$. If $d \geq q-1$, then the probability generating function for $X_{n,d}$ is given by*

$$\Phi_{n,d}(t) := \sum_{k \geq 0} P(X_{n,d} = k)t^k = \left(1 + \frac{t-1}{q}\right)^{q^n}.$$

Corollary 1.2. *Under the same assumption, the expected value E and the variance V of $X_{n,d}$ are*

$$E(X_{n,d}) = q^{n-1}, \qquad V(X_{n,d}) = q^{n-1}\left(1 - \frac{1}{q}\right).$$

We obtain the following result when $n = 1$ by taking the limit $q \to \infty$.

Corollary 1.3. *If $q \to \infty$, the probability distribution of $X_{1,d}$ converges to the Poisson distribution with parameter 1.*

In Section 2, we compute the average number of zeros using the incidence variety method. Though the computation is not later used to calculate $\Phi_{n,d}(t)$, we included the proof because it 1) explains the anticipated expected value without complex calculation, and 2) is flexible, as it works for many variations of the sample space, and only requires a nonnegative degree d. See Remark 2.4. Section 3 is devoted to computing $\Phi_{1,d}(t)$. Our method is based on a similar computation by Leontév [Leo06], who studied the case of single-variable monic polynomials of a fixed degree. Our result (Theorem 3.3) for $\Phi_{1,d}(t)$ is a careful application of his result.

Our main contribution is in Section 4. By employing elementary algebraic tools such as Lagrange interpolation and the Chinese remainder theorem, we show that $\Phi_{n,d}(t)$ is stable when $d \geq q - 1$, then calculate it explicitly in Theorem 4.5. Finally, in the last section, we discuss some related questions.

For real polynomials, there have been numerous results on an analogous problem. If the coefficients are independent and normally distributed, a classical result of Kac asserts that the average converges to $\frac{2}{\pi} \log d$ [Kac43]. See [EK95] for a nice survey, its connection with geometry, and references. For more recent results, see, for example, [DPSZ02, NV21].

For a similar question over p-adic fields, there have also been several results, including [Eva06, BCFG22]. Since the ring of p-adic integers can be obtained by taking the inverse limit of $\mathbb{Z}/p^k\mathbb{Z}$, linking these results with the case of $\mathbb{Z}/p^k\mathbb{Z}$-coefficients will be interesting. See Question 5.9.

2. THE AVERAGE NUMBER OF ZEROS

In this section, we fix the notation. Then, using the incidence variety, we show that the average number of zeros of a random polynomial with nonnegative degree is q^{n-1}.

Let $\mathbb{F}_q$ be the finite field of order $q = p^r$, where p is a prime number. The ring of polynomials with n variables and $\mathbb{F}_q$-coefficients is denoted by $\mathbb{F}_q[x_1, \cdots, x_n]$.

Definition 2.1. Fix an integer $d \geq 0$. The subset of $\mathbb{F}_q[x_1, \cdots, x_n]$ of polynomials whose degree is at most d with respect to x_i for each $1 \leq i \leq n$ is denoted by $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$. The subset of polynomials whose total degree is at most d is denoted by $\mathbb{F}_q[x_1, \cdots, x_n]_d$.

Clearly, $\mathbb{F}_q[x_1, \cdots, x_n]_d \subset \mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$. If $n = 1$, $\mathbb{F}_q[x_1]_d = \mathbb{F}_q[x_1]_{(d)}$.

Since $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$ and $\mathbb{F}_q[x_1, \cdots, x_n]_d$ are finite-dimensional $\mathbb{F}_q$-vector spaces, we can choose a random polynomial by fixing a basis and uniformly selecting coefficients.

Definition 2.2. A point $\mathbf{p} := (a_1, \cdots, a_n) \in \mathbb{F}_q^n$ is a *zero* of a polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]$ if $f(\mathbf{p}) = 0$.

Thus, we only consider zeros lying in $\mathbb{F}_q^n$.

Theorem 2.3. *Fix $d \geq 0$. The average number of zeros of a random polynomial in $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$ is q^{n-1}.*

Proof. Consider the *incidence variety*

$$\mathrm{Inc} := \{(\mathbf{p}, f) \mid f(\mathbf{p}) = 0\} \subset \mathbb{F}_q^n \times \mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}.$$

Then, there are two projection maps

$$
\begin{array}{ccc}
 & \mathrm{Inc} & \\
\pi_1 \swarrow & & \searrow \pi_2 \\
\mathbb{F}_q^n & & \mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)},
\end{array}
$$

where $\pi_1(\mathbf{p}, f) := \mathbf{p}$ and $\pi_2(\mathbf{p}, f) := f$. For each $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$, $|\pi_2^{-1}(f)|$ is the number of zeros of f. Thus, the desired average is given by

$$\frac{|\mathrm{Inc}|}{|\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}|}.$$

First of all, the dimension of $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$ is $(d+1)^n$. On the other hand, for $|\mathrm{Inc}|$, observe that, for any $\mathbf{p} \in \mathbb{F}_q^n$, we may define the evaluation map $\mathrm{ev}_{\mathbf{p}} : \mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)} \to \mathbb{F}_q$ by setting $\mathrm{ev}_{\mathbf{p}}(f) := f(\mathbf{p})$. Then, for each $\mathbf{p} \in \mathbb{F}_q^n$, $\pi_1^{-1}(\mathbf{p})$ is in bijection with $\ker \mathrm{ev}_{\mathbf{p}}$. Since $\mathrm{ev}_{\mathbf{p}}$ is clearly surjective, by the dimension theorem,

$$|\pi_1^{-1}(\mathbf{p})| = |\ker \mathrm{ev}_{\mathbf{p}}| = q^{(d+1)^n - 1}.$$

Since every fiber of π_1 has the same cardinality, for any $\mathbf{p} \in \mathbb{F}_q^n$, $|\mathrm{Inc}| = |\mathbb{F}_q^n||\pi_1^{-1}(\mathbf{p})| = q^{(d+1)^n + n - 1}$.

In sum, we have

$$\frac{|\mathrm{Inc}|}{\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}} = \frac{q^{(d+1)^n + n - 1}}{q^{(d+1)^n}} = q^{n-1}.$$

$\square$

Remark 2.4.   (1) By similar logic, it is straightforward to show that the average num-
        ber of zeros of a random polynomial in $\mathbb{F}_q[x_1, \cdots, x_n]_d$ is also q^{n-1}.
     (2) Unlike Theorem 1.1, the result does not necessarily require $d \geq q - 1$.

Since the average in Theorem 2.3 does not depend on the degree d, we can make the
following conclusion by taking the limit $d \to \infty$.

Corollary 2.5. *The average number of distinct zeros of a randomly chosen polynomial in $\mathbb{F}_q[x_1, \cdots, x_n]$*
is q^{n-1}. In particular, on average, a random polynomial in $\mathbb{F}_q[x]$ has precisely one zero.

3. SINGLE VARIABLE CASE

Here, we compute the probability generating function for the $n = 1$ variable case.

Definition 3.1. Let $X_{n,d} : \mathbb{F}_d[x_1, \cdots, x_n]_{(d, \cdots, d)} \to \mathbb{N}$ be the random variable of the number
of distinct zeros in $\mathbb{F}_q$ of a random polynomial.

The distribution of the number of zeros of a random polynomial was studied by Leontév
under a slightly different setup [Leo06] for $n = 1$. We deduce our generating function

(1) $$\Phi_{1,d}(t) := \sum_{k \geq 0} P(X_{n,d} = k)t^k$$

from his combinatorial computation. We emphasize that the $d \geq q - 1$ assumption is
unnecessary for the proof we provide here for $n = 1$ variables.

Let $B_{d,q,k}$ be the number of degree d *monic* polynomials in $\mathbb{F}_q[x]$ with k distinct zeros.
Using the inclusion-exclusion principle and residue calculation, Leontév showed that the
set of $B_{d,q,k}$ satisfies the following generating function [Leo06, Lemma 4]:

(2) $$\sum_{k \geq 0} \frac{B_{d,q,k}}{q^k} t^k = \sum_{i=0}^{d} \frac{\binom{q}{i}}{q^i}(t-1)^i.$$

From this result, we compute the probability generating function over $\mathbb{F}_q[x]_d$. We use
the convention that the zero polynomial has degree $-\infty$.

Lemma 3.2. *Let $C_{d,q,k}$ be the number of polynomials in $\mathbb{F}_q[x]_d$ with k distinct zeros. Then*

$$C_{d,q,k} = \begin{cases} (q-1)\sum_{e=0}^{d} B_{e,q,k}, & k \neq q, \\ (q-1)\sum_{e=0}^{d} B_{e,q,q} + 1, & k = q. \end{cases}$$

Proof. Recall that $\mathbb{F}_q[x]_d$ is the vector space of all polynomials of degree at most d. As
nonzero scalar multiplication does not change the number of zeros of a given polynomial,
to count the number of all polynomials with degree at most d, we obtain the first formula.
When $k = q$, we add one to account for the zero polynomial.   □

Theorem 3.3. *The probability generating function $\Phi_{1,d}(t)$ in (1) satisfies*

$$(3) \qquad \Phi_{1,d}(t) = \sum_{i=0}^{d} \frac{(q^{d+1-i}-1)\binom{q}{i}}{q^{d+1}}(t-1)^i + \frac{t^q}{q^{d+1}}.$$

Proof. Note that

$$\Phi_{1,d}(t) = \sum_{k\geq 0} P(X_{1,d}=k)x^k = \sum_{k\geq 0} \frac{C_{d,q,k}}{q^{d+1}}t^k.$$

By Lemma 3.2,

$$\Phi_{1,d}(t) = \sum_{k\geq 0}(q-1)\sum_{e=0}^{d}\frac{B_{e,q,k}}{q^{d+1}}t^k + \frac{t^q}{q^{d+1}}$$

$$(4) \qquad = \frac{q-1}{q}\sum_{k\geq 0}\sum_{e=0}^{d}\frac{B_{e,q,k}}{q^d}t^k + \frac{t^q}{q^{d+1}} = \frac{q-1}{q}\sum_{e=0}^{d}\sum_{k\geq 0}\frac{B_{e,q,k}}{q^{d-e}q^e}t^k + \frac{t^q}{q^{d+1}}$$

$$= \frac{q-1}{q}\sum_{e=0}^{d}\frac{1}{q^{d-e}}\left(\sum_{k\geq 0}\frac{B_{e,q,k}}{q^e}t^k\right) + \frac{t^q}{q^{d+1}} = \frac{q-1}{q^{d+1}}\sum_{e=0}^{d}q^e\left(\sum_{k\geq 0}\frac{B_{e,q,k}}{q^e}t^k\right) + \frac{t^q}{q^{d+1}}.$$

Using (2), we have

$$\Phi_{1,d}(t) = \frac{q-1}{q^{d+1}}\sum_{e=0}^{d}q^e\left(\sum_{i=0}^{e}\frac{\binom{q}{i}}{q^i}(t-1)^i\right) + \frac{t^q}{q^{d+1}} = \frac{q-1}{q^{d+1}}\sum_{e=0}^{d}\sum_{i=0}^{e}\frac{\binom{q}{i}}{q^i}(t-1)^i q^e + \frac{t^q}{q^{d+1}}$$

$$= \frac{q-1}{q^{d+1}}\sum_{i=0}^{d}\left(\sum_{e=i}^{d}q^e\right)\frac{\binom{q}{i}}{q^i}(t-1)^i + \frac{t^q}{q^{d+1}} = \frac{q-1}{q^{d+1}}\sum_{i=0}^{d}\frac{q^i(q^{d+1-i}-1)}{q-1}\frac{\binom{q}{i}}{q^i}(t-1)^i + \frac{t^q}{q^{d+1}}$$

$$= \sum_{i=0}^{d}\frac{(q^{d+1-i}-1)\binom{q}{i}}{q^{d+1}}(t-1)^i + \frac{t^q}{q^{d+1}}.$$

$\square$

From the generating function, we immediately obtain the mean and the variance of the number of zeros.

Corollary 3.4. *Let E and V respectively denote the mean and variance of $X_{1,d}$. Then*

$$(5) \qquad E(X_{1,d}) = 1, \qquad V(X_{1,d}) = 1 - \frac{1}{q}.$$

Proof. Since $E(X_{1,d}) = \sum_{k\geq 0} kP(X_{1,d}=k) = \Phi'_{1,d}(1)$, and

$$V(X_{1,d}) = E(X_{1,d}^2) - E(X_{1,d})^2 = \sum_{k\geq 0}k^2 P(X_{1,d}=k) - \left(\sum_{k\geq 0}kP(X_{1,d}=k)\right)^2$$

$$= \Phi''_{1,d}(1) + \Phi'_{1,d}(1) - (\Phi'_{1,d}(1))^2,$$

the result immediately follows from Theorem 3.3 by computing the first and second derivatives of $\Phi_{1,d}(t)$ at the point $t = 1$. □

When d is large relative to q, the generating function $\Phi_{1,d}(t)$ reduces to a surprisingly simple formula.

Corollary 3.5. *If $d \geq q - 1$, the formula in (3) simplifies to*

$$\Phi_{1,d}(t) = \left(1 + \frac{t - 1}{q} \right)^q.$$

Proof. When $d \geq q$, since $P(X_{1,d} = k) = 0$ for $k > q$, we have

$$\Phi_{1,d}(t) = \sum_{i=0}^{q} \frac{(q^{d+1-i} - 1)\binom{q}{i}}{q^{d+1}}(t - 1)^i + \frac{t^q}{q^{d+1}}.$$

Using the binomial theorem, we obtain

$$\Phi_{1,d}(t) = \sum_{i=0}^{q} \binom{q}{i} \left(\frac{t - 1}{q} \right)^i - \sum_{i=0}^{q} \frac{\binom{q}{i}}{q^{d+1}}(t - 1)^i + \frac{t^q}{q^{d+1}}$$

$$= \left(1 + \frac{t - 1}{q} \right)^q - \frac{(1 + (t - 1))^q}{q^{d+1}} + \frac{t^q}{q^{d+1}} = \left(1 + \frac{t - 1}{q} \right)^q.$$

Now, a routine calculation shows that $\Phi_{1,q-1}(t) = \Phi_{1,q}(t)$, yielding the desired result. □

Corollary 3.5 implies the following interesting consequence, generalizing what Leontév observed for monic polynomials of degree $d = q - 1$.

Corollary 3.6. *For $d \geq q - 1$, as $q \to \infty$, the probability distribution of the number of zeros of a random polynomial in $\mathbb{F}_q[x]_d$ converges to a Poisson distribution with parameter 1.*

Proof. The result follows immediately from

$$\lim_{q \to \infty} \Phi_{1,d}(t) = \lim_{q \to \infty} \left(1 + \frac{t - 1}{q} \right)^q = e^{t-1}.$$

□

One may wonder why the probability distribution does not change if $d \geq q - 1$. In the next section, we explain why this is the case in a more general context.

4. MULTIVARIABLE CASE

Next, we extend the computation in Section 3 to the general case. We additionally assume that $d \geq q - 1$.

Recall that the multiplicative group $\mathbb{F}_q^*$ of nonzero elements in $\mathbb{F}_q$ is cyclic [Hun80, Theorem V.5.3]. Thus, for every $a \in \mathbb{F}_q$, we have $a^q = a$. This implies that every polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]$ in the ideal

$$(6) \qquad I := (x_1^q - x_1, \cdots, x_n^q - x_n)$$

satisfies $f(\mathbf{p}) = 0$ for all $\mathbf{p} \in \mathbb{F}_q^n$.

The following observation is essentially the Lagrange interpolation. For all $a \in \mathbb{F}_q$, let

$$(7) \qquad L_a(x) := \prod_{b \neq a} \frac{(x - b)}{(a - b)}.$$

Note that $L_a(x)$ has degree $q - 1$, $L_a(a) = 1$, and $L_a(b) = 0$ for all $b \in \mathbb{F}_q \setminus \{a\}$.

Lemma 4.1. *Every $f \in \mathbb{F}_q[x_1, ..., x_n]$ can be written as*

$$(8) \qquad f = \sum_{a \in \mathbb{F}_q} L_a(x_n) f(x_1, \cdots, x_{n-1}, a) + \sum_{i=1}^n (x_i^q - x_i) h_i$$

for some $h_1, \cdots, h_n \in \mathbb{F}_q[x_1, ..., x_n]$, where h_i is not a multiple of $(x_j^q - x_j)$ for $j < i$.

Proof. It is straightforward to check that, for every $\mathbf{p} = (p_1, \cdots, p_n) \in \mathbb{F}_q^n$,

$$f(\mathbf{p}) = \sum_{a \in \mathbb{F}_q} L_a(p_n) f(p_1, \cdots, p_{n-1}, a).$$

Thus, $f - \sum_{a \in \mathbb{F}_q} L_a(x_n) f(x_1, \cdots, x_{n-1}, a) \in I$ in (6), implying it is an $\mathbb{F}_q[x_1, \cdots, x_n]$-linear combination of $\{x_i^p - x_i\}$. That is, we can find $h_1, \cdots, h_n$. Finally, we may impose the last condition by rearranging the coefficients if h_i divides $x_j^q - x_j$ for some $j < i$. $\qquad \square$

Lemma 4.2. *For any q polynomials $\{g_a\}_{a \in \mathbb{F}_q} \subset \mathbb{F}_q[x_1, \cdots, x_{n-1}]_{(q-1, \cdots, q-1)}$, there exists a unique $f \in \mathbb{F}_q[x_1, ..., x_n]_{(q-1, \cdots, q-1)}$ such that $f(x_1, ..., x_{n-1}, a) = g_a(x_1, \cdots, x_{n-1})$ for all $a \in \mathbb{F}_q$.*

Proof. For each $a \in \mathbb{F}_q$, consider the system of congruences

$$(9) \qquad f \equiv g_a \mod (x_n - a).$$

Since the principal ideals $(x_n - a)$ are pairwisely relatively prime, we may apply the Chinese remainder theorem [Hun80, Theorem III.2.25]. By the theorem, there exists a unique $f \in \mathbb{F}_q[x_1, \cdots, x_n]$ modulo the ideal

$$\bigcap_{a \in \mathbb{F}_q} (x_n - a) = (\prod_{a \in \mathbb{F}_q} (x_n - a)) = (x_n^q - x_n)$$

satisfying the congruence relations in (9). Indeed, by using formula (8) in Lemma 4.1, we can construct such an $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(q-1, \cdots, q-1)}$ explicitly. $\qquad \square$

Proposition 4.3. *Suppose that $d \geq q - 1$. Then, the probability distribution of the number of zeros of a random polynomial in $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$ is independent of d.*

Proof. Combining Lemma 4.1 and Lemma 4.2, a polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$ can be uniquely chosen by selecting the polynomials $\{g_a\}_{a \in \mathbb{F}_q} \subset \mathbb{F}_q[x_1, \cdots, x_n]_{(q-1,\cdots,q-1)}$ and $h_1, \cdots, h_n \in \mathbb{F}_q[x_1, \cdots, x_n]$ with certain divisibility and degree conditions. Note that the choice of $h_1, \cdots, h_n$ does not affect the number of zeros of f because they are multiplied with $x_i^q - x_i$. Therefore, while computing the distribution of the number of zeros, we may assume that the sample space is $\mathbb{F}_q[x_1, \cdots, x_n]_{(q-1,\cdots,q-1)}$. $\qquad\square$

Remark 4.4.   (1) Equivalently, we may assume that the sample space is

$$(\mathbb{F}_q[x_1, \cdots, x_{n-1}]_{(q-1,\cdots,q-1)})^q.$$

This has an important consequence – the independence of the choice of $\{g_a\}_{a \in \mathbb{F}_q}$.
  (2) Proposition 4.3 explains why the probability generating function in Corollary 3.5 stabilizes.

For the proof, we henceforth assume that $d = q - 1$. For $d \geq q$, we may obtain the same result by applying Proposition 4.3.

Theorem 4.5. *The probability generating function*

$$\Phi_{n,d}(t) = \sum_{k \geq 0} P(X_{n,d} = k) t^k$$

of the number of zeros of a random polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$ is given by

$$\Phi_{n,d}(t) = (\Phi_{1,d}(t))^{q^{n-1}} = \left(1 + \frac{t-1}{q}\right)^{q^n}.$$

Proof. We proceed by induction on n. The $n = 1$ case holds by Corollary 3.5.

For a random polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$, let $N(f)$ denote the number of zeros of f. Then, observe that choosing f is equivalent to choosing q random polynomials $\{g_a\}_{a \in \mathbb{F}_q} \subset \mathbb{F}_q[x_1, \cdots, x_{n-1}]_{(d,\cdots,d)}$, where each $g_a = f(x_1, \cdots, x_{n-1}, a)$. By (8) (with $h_i = 0$), $N(f) = \sum_{a \in \mathbb{F}_q} N(g_a)$. Then, we have

$$(10) \qquad P(N(f) = k) = \sum_{\sum_{a \in \mathbb{F}_q} k_a = k} P(N(g_a) = k_a, \ \forall a \in \mathbb{F}_q).$$

Since the choice of $\{g_a\}_{a \in \mathbb{F}_q}$ are independent, the right-hand side of (10) equals

$$\sum_{\sum_{a \in \mathbb{F}_q} k_a = k} \left(\prod_{a \in \mathbb{F}_q} P(N(g_a) = k_a) \right).$$

We define a multivariable generating function

$$(11) \qquad \Phi(t_a, \ a \in \mathbb{F}_q) := \sum_{\sum_{a \in \mathbb{F}_q} k_a = k} \left(\prod_{a \in \mathbb{F}_q} P(N(g_a) = k_a) t_a^{k_a} \right).$$

The right-hand side of (11) is given by the Cauchy product of the coefficients of polynomials. Therefore,

$$(12) \quad \Phi(t_a, \ a \in \mathbb{F}_q) = \prod_{a \in \mathbb{F}_q} \left(\sum_{k_a \geq 0} P(N(g_a) = k_a) t_a^{k_a} \right) = \prod_{a \in \mathbb{F}_q} \Phi_{n-1,d}(t_a) = \prod_{a \in \mathbb{F}_q} (\Phi_{1,d}(t_a))^{q^{n-2}}$$

by the induction hypothesis.

On the other hand, setting $t_a = t$ for all $a \in \mathbb{F}_q$, we have

$$\Phi(t, \cdots, t) = \sum_{\{k_a\}_{a \in \mathbb{F}_q}} \prod_{a \in \mathbb{F}_q} P(N(g_a) = k_a) t^{\sum k_a}$$

$$(13)$$

$$= \sum_{k \geq 0} \sum_{\sum_{a \in \mathbb{F}_q} k_a = k} \left(\prod_{a \in \mathbb{F}_q} P(N(g_a) = k_a) \right) t^k = \sum_{k \geq 0} P(N(f) = k) t^k = \Phi_{n,d}(t).$$

Then, combining (12) and (13),

$$\Phi_{n,d}(t) = \Phi(t, \cdots, t) = \left(\Phi_{1,d}(t))^{q^{n-2}} \right)^q = (\Phi_{1,d}(t))^{q^{n-1}} = \left(1 + \frac{t-1}{q} \right)^{q^n}.$$

$\square$

As in the single variable case, we may compute the mean and variance from the generating function $\Phi_{n,d}(t)$. We leave the details of the computation to the interested readers.

Corollary 4.6. *The mean E and variance V of $X_{n,d}$ are given by*

$$E(X_{n,d}) = q^{n-1}, \qquad V(X_{n,d}) = q^{n-1} \left(1 - \frac{1}{q} \right).$$

5. QUESTIONS

We leave a few related questions, supported by numerical computations.

In the multivariable case, as shown in Section 4, choosing $\mathbb{F}_q[x_1, \cdots, x_n]_{(d, \cdots, d)}$ enables us to use induction on the number of variables. Perhaps another natural choice of a finite sample space is $\mathbb{F}_q[x_1, \cdots, x_n]_d$, the set of polynomials whose total degree is at most d. However, a similar argument does not work for $\mathbb{F}_q[x_1, \cdots, x_n]_d$, as the numbers of zeros for the restrictions $f(x_1, \cdots, x_{n-1}, a)$ and $f(x_1, \cdots, x_{n-1}, a')$ are not independent.

Question 5.1. What is the probability generating function of the number of distinct zeros of a random polynomial in $\mathbb{F}_q[x_1, \cdots, x_n]_d$?

From numerical investigation, it seems that the probability distribution for $\mathbb{F}_q[x_1, \cdots, x_n]_d$ resembles the distribution for $\mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$. Note that this is not entirely obvious, as the ratio of the dimensions of $\mathbb{F}_q[x_1, \cdots, x_n]_d$ and $\mathbb{F}_q[x_1, \cdots, x_n]_{(d,\cdots,d)}$ is $\binom{n+d-1}{d}/(d+1)^n$, which approaches zero as $n \to \infty$. On the other hand, from Remark 2.4, we know their expected values are the same.

Another natural extension is the case of polynomials over a projective space. More precisely, let $\mathbb{F}_q[x_1, \cdots, x_n]_d^h$ be the $\mathbb{F}_q$-vector space consisting of homogeneous polynomials of degree d. Let $Y_{n,d}$ be the random variable of the number of nontrivial zeros up to nonzero scalar multiplication of a random polynomial $f \in \mathbb{F}_q[x_1, \cdots, x_n]_d^h$. By employing the incidence variety method in Section 2, it is straightforward to check that the expected value of $Y_{n,d}$ is

$$E(Y_{n,d}) = \frac{(q^{n+1} - 1)\left(q^{\binom{n+d}{d}-1} - 1\right)}{(q-1)\left(q^{\binom{n+d}{d}} - 1\right)} \approx q^{n-1}.$$

Question 5.2. What is the probability generating function of $Y_{n,d}$?

Another possible line of inquiry is the distribution for a random polynomial over $\mathbb{Z}/m\mathbb{Z}$, where m may be composite. In particular, let $K_{n,d}^m$ be the random variable of the number of distinct zeros of a random polynomial in $\mathbb{Z}/m\mathbb{Z}[x_1, \cdots, x_n]_{(d,\cdots,d)}$.

Question 5.3. Suppose $m = p_1 \cdots p_r$ is the square-free product of r distinct primes. What is the probability generating function of $K_{n,d}^m$?

Remark 5.4. By the Chinese remainder theorem, any random $f \in \mathbb{Z}/m\mathbb{Z}[x_1, \cdots, x_n]_{(d,\cdots,d)}$ is equivalent to a random r-tuple of polynomials $(f_{p_1}, \cdots, f_{p_r})$, where each f_{p_i} denotes the reduction of f in $\mathbb{Z}/p_i\mathbb{Z}[x_1, \cdots, x_n]_{(d,\cdots,d)}$.

Despite this fact, computing the probability generating function of $K_{n,d}^m$ remains a non-trivial task. Numerically, however, one can discern some statistical patterns.

Conjecture 5.5. *If m is square-free, then the expected value E of $K_{n,d}^m$ is*

$$E(K_{n,d}^m) = m^{n-1}.$$

Further, the distribution is stable for $d \geq p_r - 1$, where p_r is the largest prime factor of m.

Another potential question is the following:

Question 5.6. If m is some prime power p^k, what is the probability generating function of $K_{n,d}^m$?

Numerically, despite similarities to Question 5.3, this case presents some additional subtleties. We record our observations for particular values of m, d and n below with a supporting graph (Figure 1).

Conjecture 5.7. *If m is some prime power p^k, then the expected value E of $K_{n,d}^m$ is also*

$$E(K_{n,d}^m) = m^{n-1}.$$

In the single variable case, for $d > 1$, the variance V of $K_{1,d}^m$ is

$$V(K_{1,d}^m) = k\left(1 - \frac{1}{p}\right).$$

Further, the distribution stabilizes as d increases.

For $n \geq 1$ variables, one can also observe the following interesting behavior regarding the probability distribution of $K_{n,d}^m$.

Observation 5.8. *If m is some prime power p^k, the probability that $K_{n,d}^m = \beta$ is nonzero only when β is a multiple of p^{n-1}.*
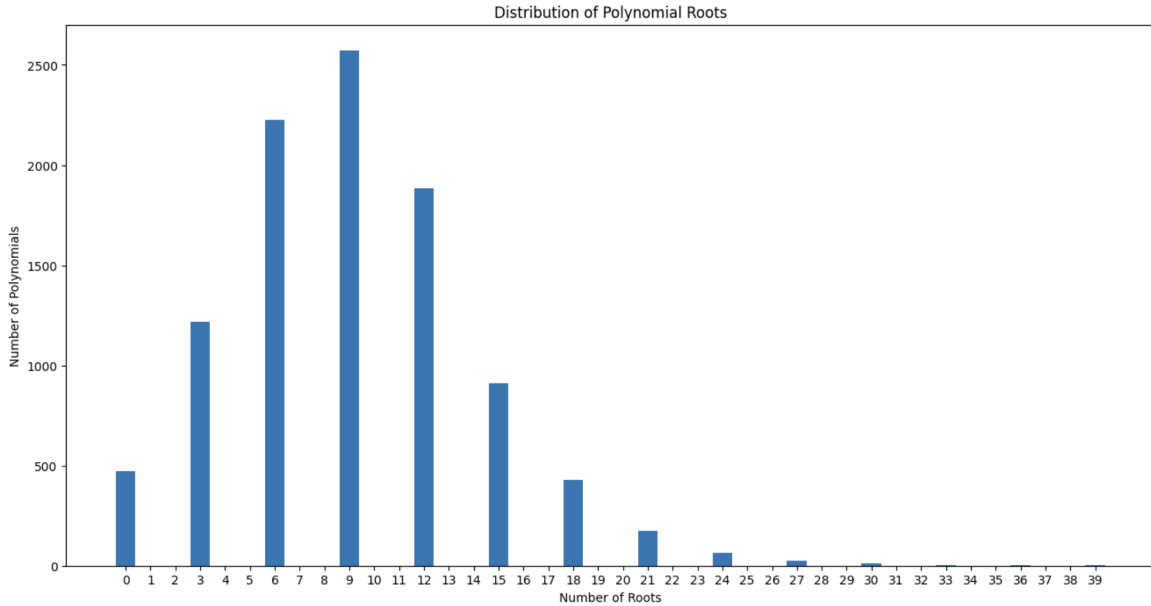

FIGURE 1. Distribution of zeros over $\mathbb{Z}/9\mathbb{Z}[x_1, x_2]_{(2,2)}$.

One might inquire similarly about polynomials with integer coefficients. However, it seems that obtaining an integral zero for a random polynomial in $\mathbb{Z}[x]$ is an exceedingly rare event. In other words, the expected number of integral zeros of a random polynomial in $\mathbb{Z}[x]_d$ is zero [BSK20] (we encourage readers to consider the $d = 1$ case by themselves).

Finally, another interesting question arises if we consider the p-adic numbers.

Question 5.9. Let $\mathbb{Z}_p$ be the set of p-adic integers and $W_{n,d}$ be the random variable of the number of zeros of a random polynomial in $\mathbb{Z}_p[x_1, \cdots, x_n]_{(d, \cdots, d)}$. What is the probability distribution of $W_{n,d}$? One may ask a similar question for $\mathbb{Q}_p$, the field of fractions of $\mathbb{Z}_p$. Can we relate this distribution to the case of $\mathbb{Z}/p^k\mathbb{Z}[x_1, \cdots, x_n]_{(d, \cdots, d)}$?

REFERENCES

[BSK20] L. Bary-Soroker and G. Kozma. Irreducible polynomials of bounded height. *Duke Math. J.* 169(2020), no.4, 579–598. 11

[BCFG22] M. Bhargava, J. Cremona, T. Fisher, and S. Gajović. The density of polynomials of degree n over $\mathbb{Z}_p$ having exactly r roots in $\mathbb{Q}_p$. *Proc. Lond. Math. Soc.* (3) 124 (2022), no. 5, 713–736. 2

[DPSZ02] A. Dembo, B. Poonen, Q.-M. Shao, and O. Zeitouni. Random polynomials having few or no real zeros *J. Amer. Math. Soc.* 15 (2002), no. 4, 857–892. 2

[EK95] A. Edelman and E. Kostlan, How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc.* 32 (1995), 1–37. 2

[Eva06] S. Evans, The expected number of zeros of a random system of p-adic polynomials. *Electron. Comm. Probab.* 11 (2006), 278–290. 2

[Kac43] M. Kac, On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* 49 (1943), 314–320 and 938. 2

[Hun80] T. Hungerford, *Algebra*. Graduate Texts in Mathematics, 73. Springer-Verlag, New York-Berlin, 1980. xxiii+502 pp. 7

[Leo06] V. K. Leontév, On the roots of random polynomials over a finite field. *Mat. Zametki* 80 (2006), no.2, 313–316; translation in *Math. Notes* 80 (2006), no.1–2, 300–304. 2, 4

[NV21] O. Nguyen and V. Vu. Random polynomials: central limit theorems for the real roots. *Duke Math. J.* 170 (2021), no. 17, 3745–3813. 2

RITIK JAIN, DEPARTMENT OF MATHEMATICS, FORDHAM UNIVERSITY, NEW YORK, NY 10023

Email address: rjain14@fordham.edu

HAN-BOM MOON, DEPARTMENT OF MATHEMATICS, FORDHAM UNIVERSITY, NEW YORK, NY 10023

Email address: hmoon8@fordham.edu

PETER WU, DEPARTMENT OF MATHEMATICS, FORDHAM UNIVERSITY, NEW YORK, NY 10023

Email address: pwu34@fordham.edu