Counting Polynomials over Finite Fields: Random Properties and Algorithms

Daniel Panario
School of Mathematics and Statistics
Carleton University
daniel@math.carleton.ca

Jo60 Conference – May 29, 2010

Introduction

Let q be a prime power. In this talk, we consider univariate polynomials over a finite field $\mathbb{F}_q$.

We are interested in the following aspects:

- counting polynomials with special forms;
- random polynomials in algorithms;
- average-case analysis of algorithms; and
- decomposition of random polynomials in its irreducible factors.

It is well-known that a polynomial of degree n over $\mathbb{F}_q$ is irreducible with probability close to $1/n$.

Can we say something more?

- How many irreducible factors a random polynomial has?
- How often will it be squarefree or k-free?
- What is the expected largest (smallest) degree among its irreducible factors?
- How is the degree distribution among its irreducible factors?
- How often a polynomial is m-smooth (all irreducible factors of degree $\leq m$)?
- How often two polynomials are m-smooth and coprime?
- How is the degree distribution among the irreducible factors of the gcd of several polynomials?
- What is the expected degree of the splitting field of a random polynomial?

and so on.

Algebraic algorithms that deal with polynomials over finite fields can often be analyzed counting polynomials with particular properties. Examples:

- irreducibility tests for polynomials,
- polynomial factorization,
- gcd computations, and
- discrete logarithm problem.

The most important characteristics of these algorithms can be treated systematically by a methodology based on generating functions and asymptotic analysis: analytic combinatorics.

This methodology relates finite fields and their applications to combinatorics and number theory.

Schedule of the talk

- Basic methodology:
  - generating functions and asymptotic analysis;
  - example: expected number of irreducible factors.
- Algorithms:
  - “folklore” polynomial factorization;
  - analysis of intervals.
- Random properties:
  - what is the degree distribution of the gcd computation of several polynomials?
  - what do random polynomials look like?
- Conclusions and related problems.

General framework

Let I_n be the number of monic irreducible polynomials in $\mathbb{F}_q$. The generating functions of monic irreducible polynomials and monic polynomials are

$$I(z) = \sum_{j\geq 1} I_j z^j, \qquad \text{and}$$

$$P(z) = \prod_{j\geq 1} (1 + z^j + z^{2j} + \cdots)^{I_j} = \prod_{j\geq 1} (1 - z^j)^{-I_j}.$$

Since $[z^n]P(z)$ is q^n, we have $P(z) = (1 - qz)^{-1}$, and these relations implicitly determine I_n

$$I_n = \frac{q^n}{n} + O\left(\frac{q^{n/2}}{n}\right).$$

A fraction very close to $1/n$ of the polynomials of degree n over $\mathbb{F}_q$ are irreducible.

From

$$\frac{1}{1-qz} = \prod_{j=1}^{\infty} (1-z^j)^{-I_j},$$

we get

$$\log \frac{1}{1-qz} \;=\; \sum_{j\geq 1} (I_j) \log(1-z^j)^{-1} = \sum_{j\geq 1} \frac{I(z^j)}{j}.$$

Expanding the log and equating the coefficients in z^n we get

$$\frac{q^n}{n} = \sum_{k|n} \frac{I_{n/k}}{k}.$$

Mobius inversion formula gives

$$I_n = \frac{1}{n} \sum_{k|n} \mu(k) q^{n/k}.$$

As usual, we consider bivariate generating functions to take care of critical parameters of the problems we are interested in. Asymptotic analysis is then used to extract coefficient information.

Example: expected number of irreducible factors. Let

$$P(u,z) = \prod_{j\geq 1} (1 + uz^j + u^2 z^{2j} + \cdots)^{I_j} = \prod_{j\geq 1} (1 - uz^j)^{-I_j}$$

where $[u^k z^n] P(u,z)$ is the number of polynomials of degree n with k irreducible factors.

Differentiating two times with respect to the parameter, putting $u = 1$ and asymptotic analysis gives expectation $\log n$ and standard deviation $\sqrt{\log n}$.

Flajolet and Soria (1990) prove that the number of irreducible factors has a Gaussian distribution.

Theorem. Let Ω_n be a random variable counting the number of irreducible factors of a random polynomial of degree n over $\mathbb{F}_q$, where each factor is counted with its order of multiplicity.

1. The mean value of Ω_n is asymptotic to $\log n + O(1)$ (Berlekamp; Knuth).
2. The variance of Ω_n is asymptotic to $\log n + O(1)$ (Knopfmacher and Knopfmacher; Flajolet and Soria).
3. For any two real constants $\lambda < \mu$,

$$\Pr\left\{\log n + \lambda\sqrt{\log n} < \Omega_n < \log n + \mu\sqrt{\log n}\right\} \to \frac{1}{\sqrt{2\pi}}\int_{\lambda}^{\mu} e^{-t^2/2}dt.$$

4. The distribution of Ω_n admits exponential tails (Flajolet and Soria).
5. A local limit theorem holds (Gao and Richmond).
6. The behaviour of $\Pr\{\Omega_n = m\}$ for all m is known (Cohen; Car; Hwang).

A general factorization algorithm

Folklore Algorithm

ERF Elimination of repeated factors replaces a polynomial by a squarefree one which contains all the irreducible factors of the original polynomial with exponents reduced to 1.

DDF Distinct-degree factorization splits a squarefree polynomial into a product of polynomials whose irreducible factors have all the same degree.

EDF Equal-degree factorization factors a polynomial whose irreducible factors have the same degree.

Analysis of intervals

One drawback of the DDF algorithm is that most of the gcds computed will be equal to 1. Since a random polynomial of degree n has about $\log n$ irreducible factors on average this is even more precise in that case.

How can we save gcd computations?

Von zur Gathen and Shoup (1992) and Kaltofen and Shoup (1995) present new algorithms for the DDF step based on a baby-step giant-step strategy:

> *Divide the interval $1, \ldots, n$ into about $\sqrt{n}$ intervals of size $\sqrt{n}$; for each interval, compute the joint product of the irreducible factors whose degree lies in that interval. Use DDF for every interval with more than one irreducible factor.*

The algorithms by von zur Gathen and Shoup (1992) and Kaltofen and Shoup (1995) split the interval $[1 \dots n]$ into about $\sqrt{n}$ pieces of size $\sqrt{n}$ each. When dealing with random polynomials, this breaking strategy is not the best possible.

The number of irreducible factors in a random polynomial of degree n tends to a Gaussian distribution with mean value $\log n$.

These $\log n$ factors are not equally distributed in the interval $[1, n]$: the expected number of irreducible factors of degree k in a random polynomial is roughly $1/k$. Thus, one expects to have more factors of lower degrees than of higher degrees.

When dealing with random polynomials, it is natural to consider partitions with growing interval sizes in order to avoid collision of irreducible factors in intervals.

Von zur Gathen and Gerhard (2002) use polynomially growing interval sizes to factor large degree random polynomials over $\mathbb{F}_2$.

These intervals have led to the million-degree factorization of Bonorden, von zur Gathen, Gerhard, Müller and Nöcker (2000).

The analysis of these algorithms involve studying the degree distribution of irreducible factors in intervals. A first step towards this analysis is in von zur Gathen, Panario and Richmond (submitted).

We provide useful information on the parameters related to partitions of the interval $[1, n]$:

- mean value and variance for the number of gcds executed;
- mean value and variance for the number of multi-factor intervals of a polynomial (intervals with more than one irreducible factor);
- mean value and variance for the number of irreducible factors of a polynomial whose degrees lie in any of its multi-factor intervals;
- mean value and variance for the total degree of irreducible factors (of a polynomial) whose degrees lie in any of the multi-factor intervals for the polynomial;

and so on.

Gcd computations

We are interested in the distribution of the common factors of several random monic polynomials over $\mathbb{F}_q$. Related results:

- The probability that several polynomials are coprime has been studied by Corteel, Savage, Wilf and Zeilberger (1998), and Reifegerste (2000).
- Drmota and Panario (2002) study pairs of coprime polynomials with the condition of being smooth.
- The continued fraction for polynomials has been studied by Knopfmacher and Knopfmacher (1988, 1991), and Friesen and Hensley (1996).
- The analysis of the Euclidean algorithm for polynomials over $\mathbb{F}_2$ is in Ma and von zur Gathen (1990), and for any finite field in Lhote (2006).

Let $\vec{n} = (n_1, n_2, \ldots, n_l)$, where $n_1, n_2, \ldots, n_l$ are the degrees of ℓ polynomials.

Consider the following random variables:

- $Z_r(\vec{n})$: number of irreducible factors in the gcd;
- $Z_d(\vec{n})$: number of distinct irreducible factors in the gcd;
- $Z_t(\vec{n})$: total degree of the gcd.

Gao and Panario (2004) show that the limiting distribution of $Z_t(\vec{n})$ is a geometric distribution, and the distributions of $Z_d(\vec{n})$ and $Z_r(\vec{n})$ are very close to Poisson distributions when $q \geq 64$:

- derive probability generating functions $F(z_1, \dots, z_\ell; u)$ for the above random variables;
- show that $F(z_1, \dots, z_\ell; u)$ is equal to
$$\frac{1}{1-qz_1} \cdots \frac{1}{1-qz_\ell} F_0(z_1, \dots, z_\ell; u)$$
where $F_0(z_1, \dots, z_\ell; u)$ depends on the random variable;
- show that $F_0(z_1, \dots, z_\ell; u)$ is analytic in $|z_i| < (1+\delta)/q$, $\delta > 0$, and that $F_0(1/q, \dots, 1/q; u) \neq 0$ for $|u| \leq 1 + \varepsilon$, and apply the next lemma.

The generating functions have dominant singularities at $z_j = 1/q$, $j = 1, \ldots, \ell$. We need transfer lemmas similar to those of Flajolet and Odlyzko (1990) but for multivariate generating functions.

Lemma. Let $G(z_1, \ldots, z_\ell; u)$ be a generating function equal to

$$\frac{1}{1-qz_1}\frac{1}{1-qz_2}\cdots\frac{1}{1-qz_\ell}F(z_1, \ldots, z_\ell; u),$$

where, for $|u| \leq 1+\varepsilon$, $F(z_1, \ldots, z_\ell; u)$ is analytic in

$$\left\{(z_1, \ldots, z_\ell) \colon |z_j| < \frac{1+\delta}{q},\ \delta > 0,\ j = 1, 2, \ldots, \ell\right\},$$

and $F(1/q, \ldots, 1/q; u) \neq 0$. Then, uniformly for $|u| \leq 1+\varepsilon$

$$\begin{aligned}[z_1^{n_1} \ldots z_\ell^{n_\ell}]G(z_1, \ldots, z_\ell; u) &= F(1/q, \ldots, 1/q; u)\, q^{n_1+\cdots+n_\ell} \\ &\quad \left(1 + O\left(\frac{1}{n_1} + \cdots + \frac{1}{n_\ell}\right)\right).\end{aligned}$$

A simplified picture of a random polynomial:

- it is irreducible with probability tending to 0 as $n \to \infty$;
- it is k-free with probability $1 - 1/q^{k-1}$;
- it has a factor of degree r with probability $1/r$ (not concent.);
- it has no linear factors with asymptotic probability ranging from 0.25 to $0.3678\ldots$ as q grows;
- it has $\log n$ irreducible factors (concentrated);
- it has $c_k n$ expected kth largest degree irreducible factor, where $c_1 = 0.62433\ldots$, $c_2 = 0.20958\ldots$, $c_3 = 0.08831\ldots$ and the remaining irreducible factors have small degree (here c_1 is Dickman-Golomb's constant);
- it has expected first and second smallest degree factors asymptotic to $e^{-\gamma} \log n$ and $e^{-\gamma} \log^2 n/2$ (not concentrated);
- it has irreducible factors of distinct degree with asymptotic probability between $0.6656\ldots$ and $e^{-\gamma} = 0.5614\ldots$ as $q \to \infty$.

Other results

We comment on a methodology for counting random polynomials over finite fields based on analytic combinatorics.

- Arratia, Barbour and Tavaré (1993) and Hansen (1993) study the joint degree distribution of the irreducible factors from a probabilistic point of view.
- Other techniques have been used. Dixon and Panario (2004) study the expected degree of the splitting field of a random polynomial over a finite field using methods related to the order of a permutation (Goh and Schmutz, 1991; Stong, 1998), and to the normal distribution of the logarithm of the order of a permutation (Erdös and Turan, 1967).
- Counting multivariate polynomials over finite fields has been treated by Carlitz and others, and more recently by von zur Gathen, Viola and Ziegler.

From the acknowledgements of my thesis:

> *Joachim von zur Gathen was influential at several levels. First, he proposed the topic for this thesis. Our numerous discussions at that time greatly increased my understanding of the area. Later, as a co-author, his meticulous way of working was inspiring. I hope some day I can state such interesting questions with such meticulous answers.*

Thank you Joachim!