Random Polynomials and Factorization Algorithms

Xavier Gourdon
INRIA Rocquencourt

October 4, 1993

[summary by Henry Crapo]

Introduction

We study the multiplicative structure of the set $\mathbb{F}_q[x]$ of polynomials with coefficients in a finite field $\mathbb{F}_q$ of q elements. The aim is to develop a unified treatment of probabilistic properties (mean, standard deviation, distribution) of the major parameters relevant to polynomial factoring algorithms. Several of the analyses are already known but put here in a common perspective, while some others appear to be new (smallest degree, largest degree, probability that the distinct degree factorization is complete).

The story for polynomials of large degree over some fixed field $\mathbb{F}_q$ goes as follows: A random polynomial has with high probability about $\log n$ irreducible factors, the distribution of the number of factors being Gaussian in the limit, with exponential tails. The asymptotic probability that it is square–free is between $1/2$ ($q = 2$) and 1 ($q = \infty$) while the degree of its square free part is $n - \mathcal{O}(1)$ on average. In other words, a random polynomial is expected to have only a very few repeated factors of very small degree. The number of factors of a fixed degree r is approximately Poisson distributed with parameter $1/r$. The degrees of the smallest and largest irreducible factors are on average about $0.5614 \log n$ and $0.6243\, n$. The distinct degree factorization completely factors a polynomial of large degree with probability that varies between 39% ($q = 2$) and 56% ($q = \infty$).

1. Basic equations

Let $\mathcal{S}$ be a class (species) of combinatorial structures. A class $\mathcal{P}$ of structures is said to be *decomposable* over $\mathcal{S}$ if every element of $\mathcal{P}$ is uniquely expressible as a multiset of elements of $\mathcal{S}$. For example, the cycle-structure of permutations is a multi-set of cycles, and the unitary polynomials over a finite field $\mathbb{F}_q$ are multisets of irreducible unitary polynomials.

Let $\mathcal{I}$ be a class of basic objects of various integral weights with $|\omega|$ denoting the weight of $\omega \in \mathcal{I}$, then

$$I(z) = \sum_{\omega \in \mathcal{I}} z^{|\omega|} = \sum_n I_n z^n,$$

where I_n is the number of objects in $\mathcal{I}$ having weight n. The corresponding generating functions for finite subsets or multisets of $\mathcal{I}$ are respectively

$$Q(z) = \prod_{\omega \in \mathcal{I}} (1 + z^{|\omega|}) = \prod_{n=1}^{\infty} (1 + z^n)^{I_n}, \qquad P(z) = \prod_{\omega \in \mathcal{I}} (1 - z^{|\omega|})^{-1} = \prod_{n=1}^{\infty} (1 - z^n)^{-I_n}.$$

Polynomials. Take $\mathcal{I}$ to be the collection of all monic irreducible polynomials over a finite field $\mathbb{F}_q$, with weight being degree. Let $P(z), Q(z)$, defined above, be the generating function of all monic polynomials, monic square-free polynomials, respectively. Since $P_n = [z^n]P(z)$ has value q^n, we have $P(z) = (1 - qz)^{-1}$, and the first relation implicitly determines I_n. Taking logarithms and applying Möbius inversion to the resulting expression, we obtain

$$I(z) = \sum_{k=1}^{\infty} \frac{\mu(k)}{k} \log \frac{1}{1 - z^k}, \qquad I_n = \frac{q^n}{n} + \mathcal{O}(\frac{q^{n/2}}{n}),$$

so that a fraction extremely close to $1/n$ of the polynomials of degree n are irreducible. These relations imply that $I(z)$ has only one dominant singularity at $z = 1/q$, around which $I(z) \sim \log[1/(1 - qz)]$. Thanks to the singular part of $I(z)$, we will be able to derive several asymptotic formulæ by singularity analysis [4].

The Euler relation $1/(1 - z) = (1 + z)(1 + z^2)(1 + z^4)\cdots$ applied to the infinite products for $P(z), Q(z)$ entails $P(z) = Q(z)\cdot Q(z^2)\cdot Q(z^4)\cdots$, an identity that corresponds to grouping repeated factors according to the binary representations of their multiplicities. Thus $Q(z) = P(z)/P(z^2)$ so that

$$Q(z) = \frac{1 - qz^2}{1 - qz} \qquad \text{and} \qquad Q_0 = 1, \quad Q_1 = q, \quad Q_n = q^{n-1}(q - 1)\ (n \geq 2).$$

The permutation model. The joint distribution of degrees in the prime decomposition of a random polynomial over $\mathbb{F}_q$ having degree n admits as a limit, as the cardinality q of the base field tends to infinity (n staying fixed!), the joint distribution of cycle lengths in random permutations of size n. This gives rise to a useful heuristic: probabilistic properties of polynomial factorizations often have a shape resembling that of corresponding properties of the cycle decomposition of permutations.

2. Number of irreducible factors

The number of monic irreducible factors of a monic polynomial in $\mathcal{P} = \mathbb{F}_q[x]$ is an additive parameter with bivariate generating function

$$P(z, u) = \prod_{p \in \mathcal{P}} (1 - uz^n)^{-I_n}.$$

From standard generating function techniques, the total number of monic irreducible factors in all monic polynomials of degree n is the coefficient $[z^n]T(z)$ where

$$T(z) = \frac{\partial}{\partial u} P(z, u)\Big|_{u=1} \quad = \quad P(z) \cdot \left(\sum_{k=1}^{\infty} I(z^k) \right).$$

Using Dirichlet convolution and singularity analysis we find the function $T(z)$ has only one dominant singularity at $z = 1/q$ around which

$$T(z) \sim \frac{1}{1 - qz} \cdot \log \frac{1}{1 - qz} \qquad (z \to \frac{1}{q}).$$

By transfer lemmas [4], the total number of monic irreducible factors in all the polynomial of degree n satisfies $T_n \sim q^n \log n$ as $n \to \infty$, so that the mean number of irreducible factors is asymptotically $\log n$. Pretty much everything is known regarding this parameter. The variance is known to be asymptotically $\log n$, and once normalized, the distribution is Gaussian in the limit [5] with exponential tails [6]. A local limit theorem also holds and results from the general theorem of Gao and Richmond [7].

3. Number of irreducible factors of fixed degree

THEOREM 1. *Let r be a positive integer. Let $\Omega_n(r)$ be the random variable counting the number of irreducible factors of degree r of a random polynomial of degree n over $\mathbb{F}_q$, each factor being counted with its order of multiplicity.*

(1) *The mean value $\mu_n(r)$ and variance $\sigma_n(r)^2$ of $\Omega_n(r)$ are asymptotically, as n tends to infinity*

$$\mu_n(r) \underset{n\to\infty}{\sim} \frac{I_r q^{-r}}{1-q^{-r}} \underset{q\to\infty}{\sim} \frac{1}{r}, \qquad \sigma_n(r)^2 \underset{n\to\infty}{\sim} \frac{I_r q^{-r}}{(1-q^{-r})^2} \underset{q\to\infty}{\sim} \frac{1}{r}.$$

(2) *For any fixed integer k,*

$$\Pr\{\Omega_n(r)=k\} \underset{n\to\infty}{\sim} (1-q^{-r})^{I_r} q^{-kr} \binom{I_r+k-1}{k} \underset{q\to\infty}{\sim} e^{-1/r}\frac{r^{-k}}{k!}.$$

The distribution of $\Omega_n(r)$ is approximately Poisson with parameter $1/r$.

The degree of the non-squarefree part of a polynomial in $\mathbb{F}_q[x]$ has order a small constant, this constant furthermore tends to zero as q goes to infinity. The parts of degree r have a size that decreases roughly geometrically (in q^{1-r}) with r.

4. Extreme degrees of irreducible factors

THEOREM 2. *The highest degree of the irreducible factors of a random polynomial of degree n over $\mathbb{F}_q$ has expectation asymptotic to Cn where C is a constant not depending on q, namely*

$$C = \int_0^\infty \left[1 - \exp\left(-\int_x^\infty \frac{e^{-t}}{t}\,dt\right)\right] dx \approx 0.62432965.$$

This constant C had already surfaced in [15] as the limit of ℓ_n/n, ℓ_n denoting the expected length of the longest cycle in a random permutation of n elements. The result is consistent, since the permutation model is the limit of the polynomial of $\mathbb{F}_q$ model as $q \to \infty$.

Numerical experimentations confirm our result. For $q = 2$ and degree 200 and 400 yield $0.62433383\ldots$ (thanks to Romberg convergence acceleration process), approximating the constant C with an error less than 10^{-5}. The proof looks like the one we find in [4] regarding the longest cycle in permutations.

THEOREM 3. *The probability that all irreducible factors of a random polynomial in $\mathbb{F}_q[x]$ of degree n have degree more than r is asymptotically, as $n \to \infty$,*

$$\prod_{j\le r}\left(1-\frac{1}{q^j}\right)^{I_j} \underset{q\to\infty}{\sim} e^{-H_r} \underset{r\to\infty}{\sim} \frac{e^{-\gamma}}{r}.$$

The expected degree should be approximately $\sum_{r=1}^{n} \frac{e^{-\gamma}}{r} \sim e^{-\gamma}\log n$, in accordance with the random permutation model.

THEOREM 4. *The smallest degree of the irreducible factors of a random polynomial of degree n over $\mathbb{F}_q$ has mean asymptotic to $e^{-\gamma}\log n$ where γ denotes Euler's constant. We have $e^{-\gamma} \approx 0.56145948$.*

The limit of our model as $q \to \infty$ corresponds to the permutation model, and our result is consistent since the shortest cycle in a random permutation of n elements is asymptotically $e^{-\gamma}\log n$ (see [15]).

5. Distinct degree factorization

We estimate the probability that the distinct degree factorization is the full factorization. This is of interest for factorization algorithms as direct methods are known to compute such a distinct degree factorization.

THEOREM 5. *The probability that the irreducible factors of a random polynomial in $\mathbb{F}_q[x]$ of degree n be all of distinct degrees is asymptotically, as $n \to \infty$,*

$$e^{-\gamma_q} = \prod_n \frac{1 + I_n q^{-n}}{(1 - q^{-n})^{-I_n}},$$

and $\gamma_q \to \gamma$ (Euler's constant) as q tends to infinity. We have the following numerical values $e^{-\gamma_2} \approx 0.3967$, $e^{-\gamma_3} \approx 0.4693$, $e^{-\gamma_4} \approx 0.4983$, $e^{-\gamma_5} \approx 0.5137$, and $e^{-\gamma} \approx 0.5614$.

6. Conclusions

A large class of parameters relative to the irreducible factor decomposition of polynomials can clearly be studied by these elementary techniques. Amongst the relevant literature, we cite the results of Mignotte and Nicolas who proved that the degree of the splitting field of a random polynomial of degree n is "almost surely" close to $e^{\log^2 n}$. See [12, 13].

Bibliography

[1] Berlekamp (Elwyn R.). – *Algebraic Coding Theory.* – Mc Graw-Hill, 1968, revised 1984 edition.

[2] Car (Mireille). – Factorisation dans $\mathbb{F}_q[x]$. *Comptes–Rendus de l'Académie des Sciences*, vol. 294 (Ser. I), 1982, pp. 147–150.

[3] Comtet (L.). – *Advanced Combinatorics.* – Reidel, Dordrecht, 1974.

[4] Flajolet (Philippe) and Odlyzko (Andrew M.). – Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, vol. 3, n° 2, 1990, pp. 216–240.

[5] Flajolet (Philippe) and Soria (Michèle). – Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory, Series A*, vol. 53, 1990, pp. 165–182.

[6] Flajolet (Philippe) and Soria (Michèle). – General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, vol. 114, 1993, pp. 159–180.

[7] Gao (Zhicheng) and Richmond (L. Bruce). – Central and local limit theorems applied to asymptotic enumerations IV: Multivariate generating functions. *Journal of Computational and Applied Mathematics*, vol. 41, 1992, pp. 177–186.

[8] Goulden (Ian P.) and Jackson (David M.). – *Combinatorial Enumeration.* – John Wiley, New York, 1983.

[9] Greene (D. H.) and Knuth (D. E.). – *Mathematics for the analysis of algorithms.* – Birkhauser, Boston, 1982, 2nd edition.

[10] Knopfmacher (John) and Knopfmacher (Arnold). – Counting irreducible factors of polynomials over a finite field. *Discrete Mathematics*, vol. 112, 1993, pp. 103–118.

[11] Knuth (Donald E.). – *The Art of Computer Programming.* – Addison-Wesley, 1981, 2nd edition, vol. 2: Seminumerical Algorithms.

[12] Mignotte (M.) and Nicolas (J.-L.). – Statistiques sur $\mathbb{F}_q[x]$. *Annales de l'Institut Henri Poincaré, Série B*, vol. XIX, n° 2, 1983, pp. 113–121.

[13] Nicolas (J.-L.). – A Gaussian law on $\mathbb{F}_q[x]$. In *Topics in Classical Number Theory, Colloquia Mathematica Societatis Janos Bolyai*, vol. 34, pp. 1127–1162. – 1981.

[14] Odlyzko (A. M.). – Asymptotic enumeration methods. – Preprint, March 1993. To appear as a chapter in the *Handbook of Combinatorics*, R. Graham, M. Grötschel and L. Lovász, ed.

[15] Shepp (L. A.) and Lloyd (S. P.). – Ordered cycle lengths in a random permutation. *Transactions of the American Mathematical Society*, vol. 121, 1966, pp. 340–357.