PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 117, Number 1, January 1993

IRREGULARITIES IN THE DISTRIBUTION OF IRREDUCIBLE POLYNOMIALS

KEVIN S. MCCURLEY

(Communicated by William Adams)

ABSTRACT. We prove that there exist monic polynomials f over $\mathrm{GF}(q)$ for which $f+g$ is reducible for all $g \in \mathrm{GF}(q)[x]$ with small degree. This is the analogue for polynomials of a result of Erdös and Rankin concerning gaps between consecutive primes.

1. INTRODUCTION

Many of the classical results on the distribution of prime numbers have analogues that describe the distribution of irreducible polynomials over a finite field $\mathrm{GF}(q)$. For example, the analogue of the prime number theorem is the statement that if $I(n)$ is the number of monic irreducible polynomials of degree n over $\mathrm{GF}(q)$, then $I(n) \sim q^n/n$ as $n \to \infty$ (throughout this paper we shall regard q as fixed and implicit constants may depend on q). This statement is in fact much easier to prove than the prime number theorem since it follows immediately from the explicit formula $I(n) = n^{-1} \sum_{d|n} \mu(d) q^{n/d}$ (see Knuth [4, Exercise 4.6.2.4]). Another example of a result in the distribution of prime numbers that has an analogue for polynomials is the prime number theorem for arithmetic progressions. The analogue of this result was proved by Artin in his dissertation [1].

It is well known that the Riemann hypothesis implies that for every integer n there exists a prime p such that $p = n + O(n^{1/2} \ln^2 n)$ (throughout this paper, we shall use ln to denote the natural logarithm, and log to denote the logarithm to the base q). This result has the following natural analogue for polynomials, which follows immediately from a result of Rhin [8, Théorème 4, p. 65].

Theorem (Rhin). *Let $f \in \mathrm{GF}(q)[x]$ be monic of degree n and $1 \le m \le n$. Then the number of monic irreducible polynomials g with $\deg(f-g) < m$ is*

$$q^m/n + O(nq^{n/2}).$$

From this we may immediately deduce the following.

Received by the editors December 29, 1990 and, in revised form, May 6, 1991.
1991 *Mathematics Subject Classification.* Primary 11T06.
This work was performed under U.S. Department of Energy contract number DE-AC04-76DP00789.

Corollary. *If n is sufficiently large and $m > n/2$, then for every monic polynomial f of degree n, there exists an irreducible polynomial g of degree n such that $\deg(f-g) < m$.*

In this paper we shall prove a result in the opposite direction from this corollary. The result that we prove is in fact a direct analogue of a result of Rankin [7], who refined a method of Erdös to prove that there exist infinitely many integers n such that for all primes p,

$$|n-p| > c \ln n \ln\ln n \frac{\ln\ln\ln\ln n}{(\ln\ln\ln n)^2},$$

where c is a constant. Our result is the following

Theorem. *For every $\varepsilon > 0$, there exist infinitely many integers n and polynomials $f \in \mathrm{GF}(q)[x]$ of degree n such that $f+g$ is reducible for all $g \in \mathrm{GF}(q)[x]$ with*

$$(1) \qquad \deg(g) \le \log n + \log\log n - 2\log\log\log n + (1-\varepsilon)\log\log\log\log n.$$

The method of proof for this result is an adaptation of the method of Erdös and Rankin to polynomials.

It is natural to ask how close to best possible our result is. In this regard we may construct a "balls into buckets" heuristic model to form a conjecture, similar to the way that Cramér did for gaps between consecutive primes. Let us define an equivalence relation on the set of monic polynomials of degree n over $\mathrm{GF}(q)$ by $f \cong g$ if and only if $\deg(f-g) < m$. The question that we wish to address is how large m needs to be in order for all of the q^{n-m} equivalence classes to contain an irreducible. We might expect that the $I(n)$ irreducibles of degree n are randomly distributed among the equivalence classes, and if this were the case, then a well-known probability result concerning the *coupon collector's problem* (see [3, pp. 234, 239]) suggests that all of the classes will be nonempty if $I(n) > (1+\varepsilon)q^{n-m}\ln(q^{n-m})$. This will occur if $m > (2+\varepsilon)\log n$ and n is sufficiently large, and we are therefore led to the following

Conjecture. *For any $\varepsilon > 0$, all $n \ge n_0(\varepsilon)$, and any monic f of degree n, there exists a polynomial g of degree at most $(2+\varepsilon)\log n$ such that $f+g$ is irreducible.*

This suggests that our theorem may be close to best possible. We note in passing that a similar but somewhat less precise conjecture was made previously by Coppersmith [2], who argued that for all n we should be able to find a polynomial g of degree $O(\log_2 n)$ such that $x^n + g(x)$ is irreducible over $\mathrm{GF}(2)$. Such irreducible polynomials are useful for calculating discrete logarithms in $\mathrm{GF}(2^n)$.

2. An elementary estimate

In the proof of our result we shall employ the following elementary estimate.

Lemma. *Let $N(k, m)$ denote the number of monic polynomials of degree k over $\mathrm{GF}(q)$ all of whose irreducible factors have degree at most m. Then for*

every $\varepsilon > 0$ *there exists a constant* c *such that*

$$\sum_{k \le n} N(k, m) \ll q^n \exp\left(-\frac{n}{m} \ln \frac{n}{m} + \frac{cn}{m} + c \ln m\right), \tag{2}$$

uniformly for $n \ge m \ge (1+\varepsilon) \log n$.

Much sharper estimates than (2) have been proved by Odlyzko [6] and Lovorn [5] using the saddle point method, but our proof is much simpler and covers a wider range of values of m relative to n.

The proof of the lemma is similar in spirit to a method used by Rankin [7] to estimate $\psi(x, y)$, which is the number of positive integers $\le x$ all of whose prime factors are $\le y$. The essential difference lies in the fact that Rankin used a Dirichlet series generating function, whereas we use the identity

$$\sum_{k=0}^{\infty} N(k, m) z^k = \prod_{k=1}^{m} \left(1 - z^k\right)^{-I(k)}, \tag{3}$$

which was also the starting point of Odlyzko in his application of the saddle point method.

Proof. We begin by remarking that the result is trivial if $m \le n \le em$, so that we may henceforth assume that $n \ge em$. For $0 < z < 1$ we have

$$\begin{aligned} \sum_{k \le n} N(k, m) &\le \sum_{k \le m} q^k + z^{-n} \sum_{m < k \le n} N(k, m) z^k \\ &\ll q^m + z^{-n} \sum_{k=0}^{\infty} N(k, m) z^k. \end{aligned} \tag{4}$$

We now choose $z = q^{-1}(n/m)^{1/m}$ and use the estimate

$$\ln\left(\sum_{k=0}^{\infty} N(k, m) z^k\right) = \sum_{k=1}^{m} -I(k) \ln(1 - z^k) \ll \sum_{k=1}^{m} \frac{(qz)^k}{k}.$$

Note that

$$\begin{aligned} \sum_{k=1}^{m} \frac{(qz)^k}{k} &\le \sum_{k \le \frac{m}{\ln(n/m)}} \frac{e}{k} + \frac{\ln(n/m)}{m} \sum_{\frac{m}{\ln(n/m)} < k \le m} (qz)^k \\ &\ll \ln m + \frac{(qz)^m \ln(n/m)}{m(qz-1)} \ll \ln m + \frac{n}{m}. \end{aligned}$$

It now suffices to show that in the indicated range,

$$q^m \le q^n \exp\left(-\frac{n}{m} \ln \frac{n}{m} + \frac{cn}{m}\right).$$

We let $n = mq^{\alpha m}$, and note that it suffices to prove that

$$m \ln q \le (1 - \alpha) m q^{\alpha m} \ln q + c q^{\alpha m}, \tag{5}$$

for $0 \le \alpha \le 1$. It is however easy to verify that the quantity on the right side of (5) is increasing in α if $c > 1$ and $0 \le \alpha \le 1$. Hence the result follows.

3. Proof of the theorem

Let α and β be positive constants, $w > q$, and define

$$u = w + \log w - 2\log\log w + \beta \log\log\log w,$$
$$z = w - 1,$$
$$y = \frac{w \log\log w}{\alpha \log w}.$$

We shall "sieve" the set S of monic polynomials of degree $\leq u$ by the irreducibles of degree $\leq w$, removing for each irreducible exactly one residue class modulo that irreducible. We begin by eliminating from S the polynomials that are divisible by an irreducible of degree between y and z. Any polynomial that remains in S either has all its irreducible factors of degree at most y or is divisible by a single irreducible of degree $\geq w$. Hence the number of survivors in S after this sieving is at most

$$\sum_{n\leq u} N(n, y) + \sum_{\substack{f \text{ irreducible} \\ w\leq \deg(f)\leq u}} \sum_{\substack{g\equiv 0 \bmod f \\ \deg(g)\leq u}} 1 \ll \sum_{n\leq u} N(n, y) + \sum_{w\leq k\leq u} I(k) q^{u-k}$$
$$\ll q^u \exp\left(-\frac{u}{y}\ln\frac{u}{y} + \frac{cu}{y} + c\ln y\right) + q^u \ln\left(\frac{u}{w}\right)$$
$$\ll q^u \exp\left(-\frac{u}{y}\ln\frac{u}{y} + \frac{cu}{y} + c\ln y\right) + \frac{q^u \log w}{w},$$

by the lemma.

We next sieve by the irreducibles of degree $\leq y$ using a greedy method. Proceeding through the irreducibles p of degree $\leq y$ in some order of nondecreasing degree, we remove the residue class modulo p that contains the largest number of survivors. If p has degree d and there are currently N survivors, then by the pigeonhole principle there must be a residue class containing at least Nq^{-d} polynomials. Hence after sieving by the polynomials of degree $\leq y$, the number of survivors will be reduced at least by the multiplicative factor

$$\prod_{d\leq y} \left(1 - q^{-d}\right)^{I(d)}.$$

Note that

$$\ln\left(\prod_{d\leq y}(1 - q^{-d})^{I(d)}\right) = \sum_{d\leq y} I(d)\ln(1 - q^{-d})$$
$$\leq \sum_{d\leq y}\left(\frac{q^d}{d} + O\left(\frac{q^{d/2}}{d}\right)\right)(-q^{-d} + O(q^{-2d}))$$
$$= -\sum_{d\leq y}\frac{1}{d} + O(1) = -\ln y + O(1).$$

Hence the number of survivors after sieving by all polynomials of degree $\leq z$ is

$$\ll \frac{1}{y}\left(q^u \exp(-\frac{u}{y}\ln\frac{u}{y}+\frac{cu}{y}+c\ln y)+\frac{q^u \log w}{w}\right). \tag{6}$$

Note that for fixed α,

$$\frac{u}{y}\sim\alpha\frac{\ln w}{\ln\ln w}$$

as $w\to\infty$. Hence

$$-\frac{u}{y}\ln\frac{u}{y}+\frac{cu}{y}+c\ln y\sim(c-\alpha)\ln w, \tag{7}$$

and it follows that if we choose $\alpha > c+2$, then for w sufficiently large, the quantity on the left of (7) will be $\leq -2\ln w$. With this choice of α it is then easy to see that the quantity (6) is

$$\ll \frac{q^w(\log\log w)^{\beta-1}}{w}.$$

Since the number of irreducibles is asymptotically q^w/w, it follows that for $\beta<1$ and w sufficiently large, the number of survivors is less than the number of irreducibles of degree w. We can therefore sieve by the irreducibles of degree w, eliminating one survivor for each such irreducible, and in so doing eliminate all survivors.

We have now shown that it is possible to choose one residue class a_p for each irreducible p of degree at most w in such a way that each of the polynomials g with $\deg(g)\leq u$ satisfies at least one of the congruences $g\equiv a_p \mod p$. By the Chinese Remainder Theorem, this is equivalent to constructing

$$a \bmod \prod_{\substack{\deg(p)\leq w\\ p \text{ irreducible}}} p$$

such that for all g with $\deg(g)\leq u$, $a+g$ will be divisible by at least one irreducible of degree $\leq w$. Moreover, we may take a to have degree at most

$$\deg \prod_{\substack{\deg(p)\leq w\\ p \text{ irreducible}}} p = \sum_{k\leq w} kI(k) \ll q^w,$$

so that $\log(\deg(a))\leq w+O(1)$, and the result follows.

Acknowledgment

The author wishes to express thanks to Dan Gordon for suggestions that improved the exposition of this paper and to D. Hayes for informing me of reference [8].

References

1. E. Artin, *Quadratische Körper im Gebiet der höhern Kongruenzen.* I, II, Math. Z. **19** (1924), 153–246; *Complete papers of Emil Artin*, Addison-Wesley, Reading, MA, 1965, pp. 1–156.

2. D. Coppersmith, *Fast evaluation of discrete logarithms in fields of characteristic two*, IEEE Trans. Inform. Theory **30** (1984), 587–594.
3. William Feller, *An introduction to probability theory and its applications*, vol. I, third ed., Wiley, New York, 1970.
4. D. E. Knuth, *The art of computer programming, Vol.* 2: *Seminumerical algorithms*, 2nd ed., Addison-Wesley, Reading, MA, 1981.
5. Renet Lovorn, Ph.D. Dissertation, Dept. of Math., Univ. of Georgia, in preparation.
6. A. M. Odlyzko, *Discrete logarithms in finite fields and their cryptographic significance*, Advances in Cryptology (Proc. of Eurocrypt 1984), Lecture Notes in Computer Science, vol. 209, Springer-Verlag, New York, 1985, pp. 224–314.
7. R. A. Rankin, *The difference between consecutive prime numbers*, Proc. London Math. Soc. **13** (1938), 242–247.
8. Georges Rhin, *Repartition modulo* 1 *dans un corps de series formelles sur un corps fini*, Dissertationes Math. (Rozprawy Mat.) **95** (1972).

DIVISION 1423, SANDIA NATIONAL LABORATORIES, ALBUQUERQUE, NEW MEXICO 87185