On the number of distinct prime divisors of polynomial sequences

Harald Kümmerle, Prof. Dr. Friedrich Roesler
Faculty of Mathematics, Technische Universität München

Abstract

Let $f \in \mathbb{Z}[X]$ be a monic, irreducible polynomial of degree 3. We assume the unproven conjecture which says that the number of primitive prime divisors in $(f(n))_{1 \leqslant n \leqslant N}$ behaves asymptotically like $N \log \deg(f)$ in order to gain intuition for more specific, but important sequences of numbers related to the conjecture. These sequences seem to converge to limits that do not depend on the choice of f. For achieving our goals, combinatorial methods are used and different models are discussed.

Introduction

The study of the frequency of prime numbers in sequences has a long tradition, main results being the prime number theorem, Dirichlet's theorem on arithmetic progressions and the theorem on Pjateckii-Shapiro sequences. This Bachelor's Thesis deals with a related but simplified problem and will unfortunately not provide rigorously proven theorems for it, but it will try to give account of the degree of complexity and of the parameters that are involved in it.
It is conjectured in [1] that for an irreducible polynomial $f \in \mathbb{Z}[X]$ of degree $\geqslant 2$,

$$\pi_f(N) := \left| \left\{ p \in \mathbb{N}; p \text{ is prime} \wedge p \mid \prod_{n=1}^{N} f(n) \right\} \right|$$

behaves asymptotically like $N \log \deg(f)$. Until now, only special cases have been treated and even these only approximatively.
Instead of attacking this very difficult problem formally, we go another way: we assume the above statement in order to prove other things revolving around the distribution of π_f. We use combinatorial models to gain insight into which underlying laws control the distribution of the primitive prime divisors, namely the (apparently existing; see "experimental data") limits of the sequences

$$(\bar{\gamma}_i)_{N \in \mathbb{N}} = \left(|\{n \in \mathbb{N}; n \leqslant N \wedge f(n) \text{ has exactly i primitive prime divisors}\}| / N \right)_{N \in \mathbb{N}}$$

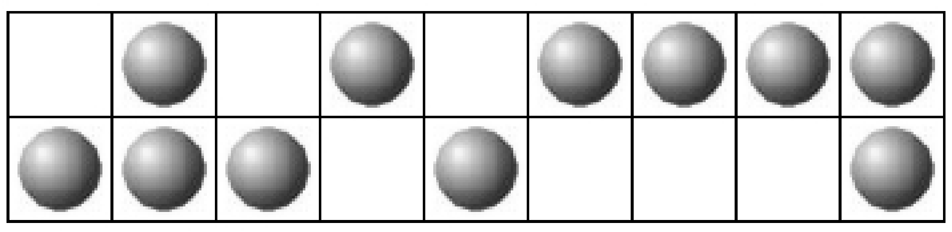
which we will call $\hat{\gamma}_i$. **(A prime p is called a primitive prime divisor (p.p.d.) of $\mathbf{a_n}$ if $\mathbf{p \mid a_n}$ and $\mathbf{p \nmid a_m}$ for $\mathbf{m < n}$.)** Furthermore, we restrict ourselves to polynomials of degree 3 for which $\hat{\gamma}_i = 0$ for all $i > 2$. All monic polynomials fulfill this.

A first Model: the "Standard Model"

The general idea of all models is that an outcome of a certain combinatorial experiment corresponds with the p.p.d.-distribution of a polynomial. A column in an outcome can be identified with an index n and its contents with the p.p.d.s of $f(n)$, respectively. For example, the p.p.d.-tableau of the polynomial $f(n) = n^3 + 2$,

n	1	2	3	4	5	6	7	8	9
$f(n)$	3	10	29	66	127	218	345	514	731
p	$\underline{3}$	$\underline{2}, \underline{5}$	$\underline{29}$	2, 3, $\underline{11}$	$\underline{127}$	2, $\underline{109}$	3, 5, $\underline{23}$	2, $\underline{257}$	$\underline{17}$, $\underline{43}$

could have been produced by the following outcome of the Standard Model that is discussed below:



A mathematical formulation of our first model is this: the Standard Model (for N columns) is the discrete probability space with

$$\left\{ S \in \mathcal{P}(\{1,2\} \times \{1,2,\ldots,N\}); |S| = \lfloor N \log 3 \rfloor \right\}$$

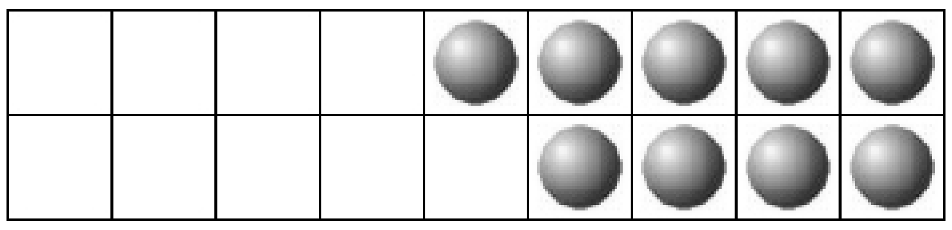
as the sample space and the equidistribution as probability measure. The choice of the cardinality of an outcome stems from our assumption that $\pi_f(N)/N \sim \log 3$.
$\lim_{N \to \infty} E^N[l_i]/N$ is the prediction for the $\hat{\gamma}_i$, where l_i is the random variable for the number of columns with exactly i elements and E^N is the expectation value function in the model with N columns.

Theorem 1. *The Standard Model predicts the following values for $\hat{\gamma}_i$:*

$$\begin{aligned} \hat{\gamma}_0 &= \left(1 - \frac{1}{2}\log 3\right)^2 \approx 0.20312, \\ \hat{\gamma}_1 &= \log(3)\left(1 - \frac{1}{2}\log(3)\right) \approx 0.49513, \\ \hat{\gamma}_2 &= \frac{1}{4}\log(3)^2 \approx 0.30173 \end{aligned}$$

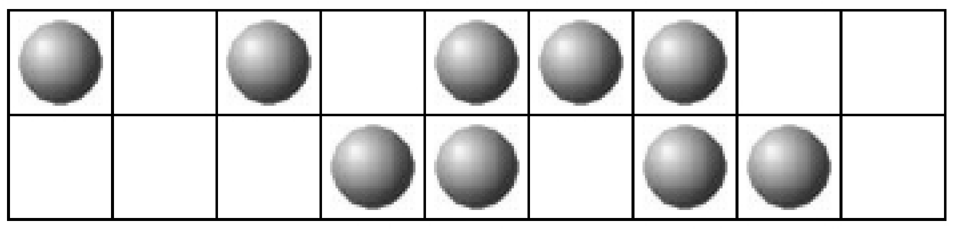
A look at the experimental data immediately indicates that further work is needed!

A refined Model: the "Compressed Model"



Outcomes like the above one cannot possibly correspond to a p.p.d.-configuration of a polynomial: the first elements of $(f(n))_{n \in \mathbb{N}}$ need to have p.p.d.s for elementary reasons. By not considering such illegetimate outcomes, the expectation values of our random variables l_i are changed and therefore the predictions for $\hat{\gamma}_i$ are altered.
The Compressed Model (for N columns) is just the Standard Model (for N columns), except that now, only outcomes are allowed whose representations fulfill the following:

To the left of the rightmost column that has at least one ball in it, every column must have at least one ball in it.



Allowed in the Standard Model, but not allowed in the Compressed Model!

A bit of interpretation is necessary to understand why this adresses the mentioned problem. This cannot be discussed here, but the numbers stand for themselves:

Theorem 2. *The Compressed Model makes the following predictions:*

$$\begin{aligned} \hat{\gamma}_0 &= \frac{12 + 8\sqrt{2} - (10 + 7\sqrt{2})\log 3}{4(3 + 2\sqrt{2})} \approx 0.06227, \\ \hat{\gamma}_1 &= \frac{4 + 3\sqrt{2}}{6 + 4\sqrt{2}} \log 3 \approx 0.77683, \\ \hat{\gamma}_2 &= \frac{2 + \sqrt{2}}{12 + 8\sqrt{2}} \log 3 \approx 0.16088 \end{aligned}$$

Lemma 3. *(One of the numerous combinatorial lemmata, [2]) Let $r \in]1; \frac{3}{2} - \frac{1}{4}\sqrt{2}[$.*

$$\sum_{i=\lfloor (r-1)N \rfloor}^{\lfloor rN/2 \rfloor} 2^{\lfloor rN \rfloor - 2i} \binom{\lfloor rN \rfloor - i}{i} \sim \frac{(1+\sqrt{2})^{\lfloor rN \rfloor + 1}}{2\sqrt{2}}.$$

Appendix: experimental data

N	$\bar{\gamma}_0$	$\bar{\gamma}_1$	$\bar{\gamma}_2$	$\bar{\gamma}_0$	$\bar{\gamma}_1$	$\bar{\gamma}_2$
100	0.05000	0.85000	0.10000	0.05000	0.73000	0.21000
500	0.05400	0.81400	0.13200	0.03400	0.79600	0.16800
1000	0.05300	0.81100	0.13600	0.03600	0.77800	0.18500
5000	0.05140	0.80860	0.14000	0.04020	0.78000	0.17920
10000	0.04940	0.81180	0.13880	0.03930	0.78250	0.17760
50000	0.05000	0.80876	0.14124	0.04116	0.78322	0.17530
100000	0.05009	0.80856	0.14135	0.04106	0.78648	0.17220
500000	0.05008	0.80844	0.14147	0.04129	0.78990	0.16863
1000000	0.05024	0.80814	0.14161	0.04161	0.79039	0.16785
5000000	0.05041	0.80767	0.14190	0.04234	0.79197	0.16558
6000000	0.05045	0.80759	0.14195	0.04235	0.79213	0.16540
7000000	0.05045	0.80753	0.14200	0.04249	0.79211	0.16530

The actual values that are obtained from the polynomials
$f(n) = n^3 + 2$ and $f(n) = 31n^3 + 53$, resp.

Thus, it seems that even non-monic irreducible polynomials behave in the same manner as monic ones. But in either case, the potential convergence is very slow. So our approximation was not bad, considering the presumed difficulty.

If they exist, then "most probably" $\hat{\gamma}_0 \approx 0.05$, $\hat{\gamma}_1 \approx 0.80$, $\hat{\gamma}_2 \approx 0.15$.

References

[1] F. Roesler. "$n^2 + 1$ revisited." *Elementare und Analytische Zahlentheorie (Tagungsband), Proceedings ELAZ-Conference May 24-28, 2004 (ed. W. Schwarz and J. Steuding)*, 267-276, Franz Steiner Verlag, 2006

[2] R. Graham, D. Knuth and O. Patashnik. "Concrete Mathematics." Addison-Wesley, sixth printing with corrections, 1990