On the cycle structure of mappings with restriction on the indegrees.

Rodrigo Martins, Daniel Panario, Eric Schmutz and Claudio Qureshi
Universidade Tecnolgica Federal do Paraná - DAMAT
rodrigomartins@utfpr.edu.br

Abstract

Let $[n] = \{1, \ldots, n\}$ and let Ω_n be the set of all functions from $[n]$ to itself, which we refer to as mappings. Let φ be random uniform element of Ω_n and let $\mathbf{T}(\varphi)$ be the order $\mathbf{T}$ of the permutation obtained by restricting φ to its cyclic nodes. B. Harris proved in 1973 an analogue for random mappings of the Erdös-Turan theorem and, in 2011, E. Schmutz obtained an asymptotic estimate on the logarithm of the expectation of $\mathbf{T}$ over all mappings on n nodes. In this work we obtain analogous results for random mappings with preimage sizes restricted to a set of the form $\{0, k\}$, for $k \geq 2$ a fixed integer. This is motivated by the use of these classes of mappings as heuristic models for the statistics of polynomials of the form $x^k + a$ over the integers modulo p, with $p \equiv 1 \pmod{k}$. We exhibit and discuss our numerical results on this heuristic.

Introduction

Let $[n] = \{1, \ldots, n\}$ and let $\varphi : [n] \to [n]$; such functions are called *mappings* in this work. The iterations of mappings has attracted interest in recent years due to applications in various areas such as physics, biology, coding theory and cryptography. For instance, Pollard's classical factorization method for integers is based on the iterations of a quadratic polynomial. The adaptation of Pollard's method to the discrete logarithm problem also relies on iterations of mappings; it is considered by some authors the best attack on the elliptic curve version of this problem.

It is known that the connected components of the directed graph associated with a mapping φ consist of a single cycle, where each cyclic node is the root of a tree directed from leaves to root. In this work we focus on asymptotic results on the cycle structure of random uniform mappings. Let $\varphi = \varphi^{(0)}$ be a mapping on n elements and consider the sequence of functional compositions $\varphi^{(m)} = \varphi \circ \varphi^{(m-1)}$, $m \geq 1$. There exists an integer $T \geq 1$ such that $\varphi^{(m+T)} = \varphi^{(m)}$ for all $m \geq n$. The least integer $\mathbf{T} = \mathbf{T}(\varphi)$ satisfying this condition equals the least common multiple of length of the cycles of φ; this is equivalent to the order of the permutation obtained by restricting the mapping f to its cyclic vertices. Erdös and Turán proved in [4] that the logarithm of this parameter over the symmetric group S_m converges in distribution to the Gaussian distribution, when centered around $\mu_m^* = \frac{1}{2}\log^2 m$ and normalized by $\sigma_m^* = \frac{1}{\sqrt{3}}\log^{3/2} m$. In 1973 Harris proved that an analogous result holds for the class of mappings under uniform distribution [5], with centralizing and normalizing constants given by $\mu_n = \frac{1}{2}\log^2 \sqrt{n}$ and normalized by $\sigma_n = \frac{1}{\sqrt{3}}\log^{3/2} \sqrt{n}$. An asymptotic estimate for the expected value of $\mathbf{T}$ over all mappings on n nodes is obtained in [9]:

$$\log \mathbb{E}_n^{\{0,k\}}[\mathbf{T}] = C_0 \frac{n^{1/3}}{\log^{2/3} n}(1 + o(1)), \tag{1}$$

where C_0 is a constant determined explicitly such that $C_0 \approx 3.36$. In this work we derive similar results for $\{0, k\}$-mappings, defined as mappings $\varphi : [n] \to [n]$ such that $|\varphi^{-1}(y)| \in \{0, k\}$ for all $y \in [n]$, where $k \geq 2$ is a fixed integer.

The research on random mappings with such restrictions is motivated in part due to the heuristic introduced by Pollard in the analysis of his factorization method [7], where one approximates the statistics of a class polynomials over finite fields by the ones of an appropriate class of mappings. This heuristic model was successfully considered by Brent and Pollard in a refined form in [3], leading to the factorization of the eighth Fermat number. For this reason this heuristic was coined as the *Brent-Pollard heuristic*.

Expected Value of T

It is known that, if $\varphi : [n] \to [n]$ is a mapping chosen uniformly at random among those satisfying $\mathbf{Z}(\varphi) = m$, then the restriction of φ to its cyclic nodes is equivalent to a random uniform permutation of the symmetric group S_m. This fact holds for random $\{0, k\}$-mappings as well [1], hence, if M_m denotes the expected order of a uniform random permutation of S_m, then we can write the expected value of $\mathbf{T}$ over all $\{0, k\}$-mapping on n nodes as

$$\mathbb{E}_n^{\{0,k\}}[\mathbf{T}] = \sum_{m=1}^{n} \mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = m] M_m. \tag{2}$$

It is clear that, if $\widetilde{m}$ is the integer that maximizes $\mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = m]M_m$ for $1 \leq m \leq n$, then

$$\mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = m_0]M_{m_0} \leq \mathbb{E}_n^{\{0,k\}}[\mathbf{T}] \leq n\mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = \widetilde{m}]M_{\widetilde{m}}, \quad \text{for all } m_0 \in [1, n]. \tag{3}$$

The quantities in the bounds above are estimated according to Lemmas 1 and 2 below.

Lemma 1 ([8])**.** *Let $k \geq 2$, $n = kh$ and $1 \leq m \leq h$. Let $\lambda = k - 1$. A random uniform $\{0, k\}$-mapping on n nodes has exactly m cyclic nodes with probability*

$$\mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = m] = \frac{\lambda m k^m}{n - m}\binom{n - m}{h - m}\binom{n}{h}^{-1}.$$

Lemma 2 ([10])**.** *Let $I = \int_0^\infty \log\log\left(\frac{e}{1-e^{-t}}\right) dt$ and $\beta_0 = \sqrt{8I}$. Then the expected order M_m of a random permutation of S_m satisfies, as m approaches infinity,*

$$\log M_m = \beta_0\sqrt{\frac{m}{\log m}} + O\left(\frac{\sqrt{m}\log\log m}{\log m}\right).$$

We extend the quantity in Lemma 1 to real numbers using the Gamma function. This allows us to define a real function $\phi_{n,\varepsilon}(x)$, for $x \in [1, n/k]$, that bounds the summand in Equation (2) by above or below, depending on the value of ϵ. We are thus able to obtain upper and lower bounds, as in Equation (3), that coincide asymptotically.

Theorem 1. *Let $k \geq 2$ be a fixed integer and, for $h \geq 1$, let $n = kh$. Let $\lambda = k - 1$. Then,*

$$\log \mathbb{E}_n^{\{0,k\}}[\mathbf{T}] = C_0 \frac{(n/\lambda)^{1/3}}{\log^{2/3}(n/\lambda)}(1 + o(1)), \quad \textit{as } h \to \infty.$$

Lognormality of T

We use once again the law of total probability according to the distribution of $\mathbf{Z}$ over $\{0, k\}$-mappings. We prove using Lemma 1 that, for each n, the distribution of $\mathbf{Z}$ over $\{0, k\}$-mappings on n nodes has a unique mode $m_\# = m_\#(n) = \sqrt{n/\lambda} + O(1)$. Let $\xi_1 = m_\#^{1-\varepsilon_n}$ and $\xi_2 = m_\#^{1+\varepsilon_n}$, where $\varepsilon_n = \log^{-3/4}(\sqrt{n/\lambda})$. We split the interval of possible values for $\mathbf{Z}$ as follows:

$$I_1 = \{m \colon 1 \leq m < \xi_1\}, \quad I_2 = \{m \colon \xi_1 \leq m \leq \xi_2\}, \quad I_3 = \{m \colon \xi_2 < m \leq n/k\}.$$

We are thus able to write $\mathbb{P}_n^{\{0,k\}}[\log \mathbf{T} \leq \mu_n + x\sigma_n] = \zeta_1 + \zeta_2 + \zeta_3$, where

$$\zeta_j = \sum_{m \in I_j} \mathbb{P}_n^{\{0,k\}}[\mathbf{Z} = m]\mathbb{P}_n^{\{0,k\}}[\log \mathbf{T} \leq \mu_n + x\sigma_n | \mathbf{Z} = m]. \tag{4}$$

An asymptotic estimate of the quantity in Lemma 1 allows us to prove that ζ_1 and ζ_3 are both $o(1)$. The convergence in distribution of $\log \mathbf{T}$ over $\{0, k\}$-mappings follows from Lemma 3 below: it is proved using use the special case $\theta = 1$ of Theorem 1.2 of [2].

Lemma 3. *Let $\mu_n = \mu_n(\lambda) = \frac{1}{2}\log^2(\sqrt{n/\lambda})$ and $\sigma_n^2 = \sigma_n^2(\lambda) = \frac{1}{3}\log^3(\sqrt{n/\lambda})$. For $m \in I_2$, let*

$$\delta_x(m, n) = \mathbb{P}_n^{\{0,k\}}\left[\frac{\log \mathbf{T} - \mu_n}{\sigma_n} \leq x \middle| \mathbf{Z} = m\right] - \phi(x),$$

and let $\Delta_x(n) = \max\{|\delta_x(m, n)|, m \in I_2\}$. Then $\Delta_x(n) = o(1)$ for any fixed $x \in \mathbb{R}$.

Theorem 2. *Let $k \geq 2$ be a fixed integer and, for $h \geq 1$, let $n = kh$. Let μ_n, σ_n^2 be as in Lemma 3. For any real number x we have*

$$\lim_{h\to\infty} \mathbb{P}_n^{\{0,k\}}[\log \mathbf{T} \leq \mu_n + x\sigma_n] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2} dt.$$

Moreover, if c is a positive constant, then the convergence is uniform for $|x| \leq c\sqrt{\log n}$.

Heuristics

In the analysis of his factorization method [7], Pollard conjectured that quadratic polynomials modulo large primes behave like random mappings with respect to their average rho length. However, it should be noted that the indegree distribution of a class of mappings impacts heavily the asymptotic distribution of a number of parameters [1]. Thus one might consider the indegree distribution of polynomials to better explore the Brent-Pollard heuristic; see the discussion in [6] for example. For instance, if f is a polynomial of the form $f(x) = x^k + a \pmod{p}$, $p \equiv 1 \pmod{k}$, then the indegree distribution of f, defined as sequence $n_j = \#\{y \in [p] \colon |f^{-1}(y)| = j\}$, $j \geq 0$, satisfies

$$n_0 = \left(1 - \frac{1}{k}\right)(p - 1), \quad n_1 = 1, \quad n_k = \frac{1}{k}(p - 1).$$

We denote this class of polynomials by $\{0, k\}$-*polynomials*. In this section we investigate the use of $\{0, k\}$-mappings as a heuristic model for $\{0, k\}$-polynomials.

Many known results on the statistics of a random mapping φ of a given class depend on its asymptotic average coalescence λ, defined as the variance of its distribution of indegrees $n_j = n_j(\varphi)$, $j \geq 0$, under uniform distribution [1]. It is worth noting that our asymptotic results on different classes of $\{0, k\}$-mappings are determined by their coalescence as well; compare Theorem 1 and Equation (1). Compare μ_n, μ_m^* and σ_n, σ_m^* as well, under the light of the fact that the expected number of cyclic nodes over all unrestricted or $\{0, k\}$-mappings are asymptotically equivalent to $\sqrt{\pi n/2}$ and $\sqrt{\pi n/2\lambda}$, respectively.

We exhibit our numerical results on the behavior of $\mathbf{T}$ over different classes of polynomials over finite fields and different classes of mappings. For $\{0, k\}$-mappings, we considered mappings on $n = p - 1$ nodes, where $p \equiv 1 \pmod{p}$. For approximately 100 primes p greater than 10^3, we consider p mappings chosen at random and all p polynomials of the form indicated in Table 1. We compute the exact value of $\mathbf{T}$ for each function and compute the corresponding average values $\overline{\mathbf{T}}(p)$. We compute the ratio $R_{\mathbf{T}}(p)$ between $\overline{\mathbf{T}}(p)$ and the quantity in Theorem 1. In Tables 1 and 2 we exhibit the average value $\overline{R_{\mathbf{T}}}$ of $R_{\mathbf{T}}(p)$ over all primes considered. The column labeled by λ indicates the asymptotic average coalescence of the corresponding class of functions.

Class of functions	λ	$\overline{R_{\mathbf{T}}}$
$\{0, 2\}$-polynomials	1	0.8031
$\{0, 3\}$-polynomials	2	0.8229
$\{0, 4\}$-polynomials	3	0.8224

Table 1: Experimental results on polynomials mod p.

Class of functions	λ	$\overline{R_{\mathbf{T}}}$
Unrestricted mappings	1	0.8090
$\{0, 2\}$-mappings	1	0.7929
$\{0, 3\}$-mappings	2	0.8304
$\{0, 4\}$-mappings	3	0.8274

Table 2: Experimental results on random mappings.

Our numerical results provide further arguments that support the Brent-Pollard heuristic: once the coalescence is taken into account according to the theoretical result of reference (Theorem 1), the relative error in the approximation of the behavior of $\mathbf{T}$ over a class of polynomials by a class of mappings with the same coalescence does not exceed 2%. This suggests that the coalescence of $\{0, k\}$-polynomials plays an prominent role in the behavior of $\mathbf{T}$ over these objects, as suggested by the Brent-Pollard heuristic.

References

[1] James Arney and Edward Bender. Random mappings with constraints on coalescence and number of origins. *Pacific Journal of Mathematics*, 103(2):269–294, 1982.

[2] Andrew D. Barbour and Simon Tavaré. A rate for the erdös-turán law. *Combinatorics, Probability and Computing*, 3(2):167–176, 1994.

[3] Richard P Brent and John M Pollard. Factorization of the eighth fermat number. *Mathematics of Computation*, 36(154):627–630, 1981.

[4] Paul Erdös and Pál Turán. On some problems of a statistical group-theory. iii. *Acta Mathematica Academiae Scientiarum Hungaricae*, 18(3-4):309–320, 1967.

[5] Bernard Harris. The asymptotic distribution of the order of elements in symmetric semigroups. *Journal of Combinatorial Theory, Series A*, 15(1):66–74, 1973.

[6] Rodrigo S V Martins and Daniel Panario. On the heuristic of approximating polynomials over finite fields by random mappings. To appear in the November Issue of the International Journal of Number Theory, 2015.

[7] John M Pollard. A monte carlo method for factorization. *BIT Numerical Mathematics*, 15(3):331–334, 1975.

[8] H Rubin and R Sitgreaves. Probability distributions related to random transformations of a finite set, 1953.

[9] Eric Schmutz. Period lengths for iterated functions. *Combinatorics, Probability and Computing*, 20(2):289–298, 2011.

[10] Richard Stong. The average order of a permutation. *Eletronic Journal of Combinatorics*, 5:R41, 1998.