TEL AVIV UNIVERSITY אוניברסיטת תל אביב

The anatomy of integers and Ewens permutations

Ofir Gorodetsky
Joint work with Dor Elboim

Horowitz Seminar, TAU
March 16, 2020

Consider S_n, the symmetric group on n elements, endowed with uniform distribution.

- π_n - a randomly uniformly drawn permutation from S_n
- $C(\pi)$ - number of cycles in π
- $C_i(\pi)$ - number of cycles in π of length i ($\sum iC_i = n$.)
- $\ell_1(\pi) \geq \ell_2(\pi) \geq \ldots$ - cycles length of π in decreasing order.

Some basic facts about permutations:

- Cauchy's formula:

$$\mathbb{P}(C_i(\pi_n) = a_i,\ i = 1, 2, \ldots, n) = \prod_{i=1}^{n} \frac{1}{(a_i)!\, i^{a_i}}.$$

- The length of the cycle containing a given element is distributed uniformly in $\{1, 2, \ldots, n\}$:

$$\mathbb{P}(\text{cycle in } \pi_n \text{ containing 1 has length } i) = \frac{1}{n}.$$

Asymptotic results about permutations:

- $\mathbb{P}(C(\pi_n) = 1) = \frac{1}{n}$.
- $\mathbb{E}C(\pi_n) = \frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n} = H_n = \log n + O(1)$.
- (Goncharov, 1941) $\mathrm{Var}C(\pi_n) \sim \log n$ and we have 'CLT':

$$\frac{C(\pi_n) - \log n}{\sqrt{\log n}} \to N(0,1).$$

- (Shepp and Lloyd, 1966)

$$\left(\frac{\ell_1(\pi_n)}{n}, \frac{\ell_2(\pi_n)}{n}, \ldots\right) \to \mathrm{PD}(1).$$

All these results have *integer* analogues.

Let N_x be an integer chosen uniformly at random from $[1, x] \cap \mathbb{Z}$.

- N_x may be factored uniquely, up to order, as a product of primes (Euclid).
  We write $N_x = p_1 p_2 p_3 \ldots p_k$, where $p_1 \geq p_2 \geq \ldots$.
- Prime factors are analogous to cycles. We set $\omega(n)$ as the number of prime factors of n (without multiplicities).

Prime Number Theorem

We have

$$\mathbb{P}(N_x \text{ is } \textit{prime}) \approx \mathbb{P}(\omega(N_x) = 1) \sim \frac{1}{\log x}.$$

Conjectured by Gauss and Legendre in 1790's. Proved by Hadamard and de la Vallée Poussin in 1896.

Distribution of ω

'Standard' computation:

$$\mathbb{E}\omega(N_x) = \sum_{p \le x} \mathbb{P}(p \mid N_x) = \sum_{p \le x} (\frac{1}{p} + O(\frac{1}{x})) \sim \log\log x.$$

Distribution of ω

'Standard' computation:

$$\mathbb{E}\omega(N_x) = \sum_{p \le x} \mathbb{P}(p \mid N_x) = \sum_{p \le x} (\frac{1}{p} + O(\frac{1}{x})) \sim \log\log x.$$

By PNT: $\sum_{p \le x} \frac{1}{p} \approx \int_2^x \frac{1}{\log t} \frac{dt}{t} \approx \log\log x$. In fact, was computed elementarily by Mertens in 1874.

Distribution of ω

'Standard' computation:

$$\mathbb{E}\omega(N_x) = \sum_{p\leq x} \mathbb{P}(p \mid N_x) = \sum_{p\leq x} (\frac{1}{p} + O(\frac{1}{x})) \sim \log\log x.$$

By PNT: $\sum_{p\leq x} \frac{1}{p} \approx \int_2^x \frac{1}{\log t}\frac{dt}{t} \approx \log\log x$. In fact, was computed elementarily by Mertens in 1874.

Theorem (Hardy-Ramanujan, 1917)

Let $g(x) \to \infty$. Then

$$\mathbb{P}(|\omega(N_x) - \log\log x| < g(x)\sqrt{\log\log x}) \to 1.$$

Distribution of ω (cont.)

Original Proof: Landau (1909) used PNT to prove

$$\mathbb{P}(\omega(N_x) = k) \sim \frac{1}{\log x} \frac{(\log \log x)^{k-1}}{(k-1)!}$$

for any fixed k. H & R proved a uniform version, at the cost of losing asymptotics:

$$\mathbb{P}(\omega(N_x) = k) \ll \frac{1}{\log x} \frac{(\log \log x + C)^{k-1}}{(k-1)!}. \qquad \square$$

Distribution of ω (cont.)

Turán's proof (1934): Computed the variance

$$\mathrm{Var}(\omega(N_x)) \sim \log\log x.$$

Now follows from Chebyshev's inequality ('second moment method'). $\square$

Distribution of ω (cont.)

Theorem (Erdös-Kac, 1940)

As $x \to \infty$,

$$\frac{\omega(N_x) - \log\log x}{\sqrt{\log\log x}} \to N(0,1).$$

Heuristic: $\omega(N_x) = \sum_p 1_{p|N_x}$. These indicators are approximately independent (at least for small p), so this is just CLT.

Largest primes

Theorem (Dickman, 1930)

The random variable $\frac{\log p_1(N_x)}{\log x}$ *has an explicit limiting distribution. The function*

$$\rho(u) = \lim_{x\to\infty} \mathbb{P}\left(p_1(N_x) \leq x^{1/u}\right)$$

is known as Dickman's ρ *function. Continuous.*

Satisfies

$$\rho(u) = \frac{1}{u}\int_{u-1}^{u} \rho(y)\,dy.$$

Appears often in complexity analysis of integer factorization algorithms.

Largest primes (cont.)

Theorem (Billingsley, 1972)

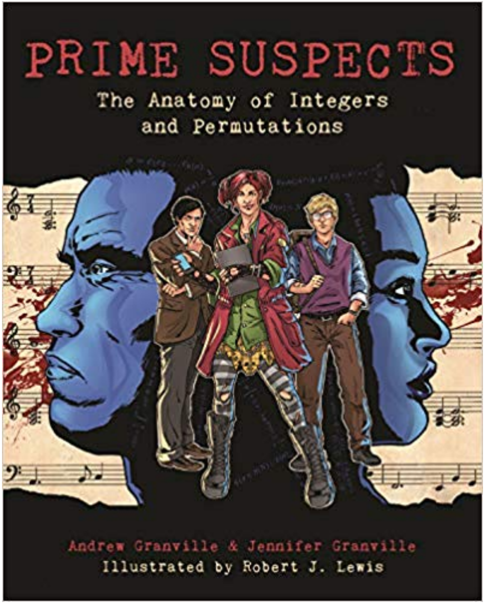
As $x \to \infty$,

$$\left(\frac{\log p_1(N_x)}{\log x}, \frac{\log p_2(N_x)}{\log x}, \ldots\right) \to PD(1).$$

A word about history: the Poisson-Dirichlet process was introduced by Kingman only in 1975. Billingsley, as well as Shepp and Lloyd's results, arrived before...

Permutation world	Integer world
$\mathbb{P}(C(\pi_n) = 1) = \frac{1}{n}$	$\mathbb{P}(\omega(N_x) = 1) \sim \frac{1}{\log x}$
$\mathbb{E}C(\pi_n) \sim \log n$	$\mathbb{E}\omega(N_x) \sim \log\log x$
$\mathrm{Var}C(\pi_n) \sim \log n$	$\mathrm{Var}\omega(N_x) \sim \log\log x$
$\frac{C(\pi_n)-\log n}{\sqrt{\log n}} \to N(0,1)$	$\frac{\omega(N_x)-\log\log x}{\sqrt{\log\log x}} \to N(0,1)$
$\left(\frac{\ell_1(\pi_n)}{n}, \ldots\right) \to PD(1)$	$\left(\frac{\log p_1(N_x)}{\log x}, \ldots\right) \to PD(1)$

Informally, permutations on n elements behave like integers of order x, where $\log x \approx n$.


PRIME SUSPECTS
The Anatomy of Integers
and Permutations
Andrew Granville & Jennifer Granville
Illustrated by Robert J. Lewis

Ewens measure with parameter $\theta > 0$: measure on S_n, defined by

$$\mathbb{P}(\pi_{n,\theta} = \pi) \propto \theta^{C(\pi)}.$$

Normalizing constant:

$$\frac{1}{n!} \sum_{\pi \in S_n} \theta^{C(\pi)} = \binom{n+\theta-1}{\theta-1} \sim \frac{n^{\theta-1}}{\Gamma(\theta)}.$$

Originally arose in population genetics (Ewens, 1972).

Chinese restaurant process

n customers enter a restaurant. Customer 1 sits at the first table. Inductively, the kth customer decides either to sit immediately to the right of one of the previous customers or to sit alone at a new table.

The probability to sit to the right of each customer is

$$\frac{1}{\theta + k - 1},$$

and the probability to open a new table is $\theta/(\theta + k - 1)$.

Exercise: the measure obtained on permutations on the customers *is Ewens*.

Some asymptotic results:

- (Hansen, 1990) $\mathbb{E}C(\pi_{n,\theta}) \sim \theta \log n$, $\mathrm{Var}C(\pi_{n,\theta}) \sim \theta \log n$ and we have 'CLT':

$$\frac{C(\pi_{n,\theta}) - \theta \log n}{\sqrt{\theta \log n}} \to N(0,1).$$

- (Watterson, 1976)

$$\left(\frac{\ell_1(\pi_{n,\theta})}{n}, \frac{\ell_2(\pi_{n,\theta})}{n}, \ldots\right) \to \mathrm{PD}(\theta).$$

The most natural analogue is

$$\mathbb{P}(N_{x,\theta} = n) \propto \theta^{\omega(n)}.$$

More generally:

$$\mathbb{P}(N_{x,f} = n) \propto f(n).$$

where f is multiplicative: $f(n \times m) = f(n) \times f(m)$ if n, m coprime.
Ubiquitous in number theory.

Examples

1. $f(n) = 1$ if n is a sum of two squares, 0 otherwise.
2. $f(n) = d(n)$, number of divisors of n.

Examples

1. $f(n) = 1$ if n is a sum of two squares, 0 otherwise.
2. $f(n) = d(n)$, number of divisors of n.
   Equivalently: pick a random pair (a, b) such that $a \cdot b \leq x$.
   Let $N_{x,f} = a \cdot b$.

Examples

1. $f(n) = 1$ if n is a sum of two squares, 0 otherwise.
2. $f(n) = d(n)$, number of divisors of n.
   Equivalently: pick a random pair (a, b) such that $a \cdot b \leq x$.
   Let $N_{x,f} = a \cdot b$.
3. $f(n) = a(n)$, the number of abelian groups of order n.

Examples

1. $f(n) = 1$ if n is a sum of two squares, 0 otherwise.
2. $f(n) = d(n)$, number of divisors of n.
   Equivalently: pick a random pair (a, b) such that $a \cdot b \leq x$.
   Let $N_{x,f} = a \cdot b$.
3. $f(n) = a(n)$, the number of abelian groups of order n.
   Equivalently: pick a random abelian group G of order $\leq n$.
   Let $N_{x,f} = |G|$.

Examples

1. $f(n) = 1$ if n is a sum of two squares, 0 otherwise.
2. $f(n) = d(n)$, number of divisors of n.
   Equivalently: pick a random pair (a, b) such that $a \cdot b \leq x$.
   Let $N_{x,f} = a \cdot b$.
3. $f(n) = a(n)$, the number of abelian groups of order n.
   Equivalently: pick a random abelian group G of order $\leq n$.
   Let $N_{x,f} = |G|$.
4. $f(n) =$ number of roots of P modulo n. (P - polynomial.)
5. $f(n) = 1$ if P has a root modulo n, 0 otherwise.

Theorem (Elboim and G., 2019)

Let f be a multiplicative function, which is on average θ on primes:

$$\frac{\sum_{p\le x} f(p)}{\sum_{p\le x} 1} = \theta + O(\log^{-A} x).$$

Then

$$\frac{\omega(N_{x,f}) - \theta \log\log x}{\sqrt{\theta \log\log x}} \to N(0,1)$$

and

$$\left(\frac{\log p_1(N_{x,f})}{\log x}, \frac{\log p_2(N_{x,f})}{\log x}, \dots\right) \to PD(\theta).$$

These results agree with properties of the Ewens measure.

Theorem (Elboim and G., 2019)

Let f be a multiplicative function, which is on average θ on primes:

$$\frac{\sum_{p\le x} f(p)}{\sum_{p\le x} 1} = \theta + O(\log^{-A} x).$$

Then

$$\frac{\omega(N_{x,f}) - \theta \log\log x}{\sqrt{\theta \log\log x}} \to N(0,1)$$

and

$$\left(\frac{\log p_1(N_{x,f})}{\log x}, \frac{\log p_2(N_{x,f})}{\log x}, \dots\right) \to PD(\theta).$$

Must require some growth condition on prime powers: $f(p^k) = O(c^k)$ for $c < \sqrt{2}$.

These results agree with properties of the Ewens measure.

Common ingredient in all of our proofs: a recent result on sums of arithmetic functions – required to get normalizing constant.

Theorem (Granville and Koukoulopoulos, 2019)

Let f be a multiplicative function, which is on average θ on primes:

$$\frac{\sum_{p\le x} f(p)}{\sum_{p\le x} 1} = \theta + O(\log^{-A} x).$$

Suppose further $f(p^k) = O(c^k)$ for $c < \sqrt{2}$. Then

$$\frac{1}{x}\sum_{n\le x} f(n) = A_f \log^{\theta-1} x + O(\log^{\theta-2} x),$$

where

$$A_f = \Gamma(\theta)^{-1} \prod_p \Big(\sum_k \alpha(p^k)/p^k\Big)(1-1/p)^{\theta}.$$

First part via Billingsley's method

Heuristically, $\alpha(p)/p$ approximates $\mathbb{P}(p \mid N_{x,f})$ in a certain range of p and x.

Truncation: replacing $(\omega(N_{x,f}) - \theta \log\log x)/(\sqrt{\theta \log\log x})$ with

$$B_x = \frac{\sum_{p \in P_x} \left(1_{p|N_{x,f}} - \frac{\alpha(p)}{p}\right)}{\sqrt{\sum_{p \in P_x} \frac{\alpha(p)}{p}(1 - \frac{\alpha(p)}{p})}}$$

where $P_x = \{\log\log\log x \leq \log\log p \leq \log\log x - \log^{1/3}\log x\}$.

Can show that $\mathbb{P}(q_1, q_2, \ldots, q_m \mid N_{x,f})$ is close to $\prod_{j=1}^{m} \alpha(q_j)/q_j$ for primes in our range.

Comparison of moments: can show that the moments of

$$C_x = \frac{\sum_{p \in P_x} \text{Bernoulli}(\frac{\alpha(p)}{p}) - \frac{\alpha(p)}{p}}{\sqrt{\sum_{p \in P_x} \frac{\alpha(p)}{p}(1 - \frac{\alpha(p)}{p})}}$$

are close to the moments of B_x.

By version of CLT, $\mathbb{E}C_x^k \to \mathbb{E}N(0,1)^k$. Hence $\mathbb{E}B_x^k \to \mathbb{E}N(0,1)^k$. Moments of normal determine distribution. □

Second part via Donnelly and Grimmett's method

Let $U_1, U_2, \ldots$ be i.i.d with distribution $\mathrm{beta}(1, \theta)$ on $[0, 1]$. Let

$$X_i = (1 - U_1) \cdots (1 - U_{i-1}) U_i.$$

Then $\{X_i\}_{i \geq 1}$ has distribution called **GEM** (Griffiths, Engen, and McCloskey). Note $\sum X_i = 1$. Also known as *stick-breaking process*.

Sorting $\{X_i\}_{i \geq 1}$ we obtain $\{Y_i\}_{i \geq 1}$ with $Y_1 \geq Y_2 \geq \ldots$. Then $\{Y_i\}_{i \geq 1}$ has **Poisson-Dirichlet** distribution with parameter θ.

Size-biased permutation of $\{Y_i\}_{i\geq 1}$: Given distinct Y_i, define $\tilde{Y}_1$ to equal Y_j with probability proportional Y_j.

Size-biased permutation of $\{Y_i\}_{i\geq 1}$: Given distinct Y_i, define $\tilde{Y}_1$ to equal Y_j with probability proportional Y_j. Inductively, $\tilde{Y}_i$ is equal Y_j (if Y_j was not chosen already) with probability proportional to Y_j.

Criterion: $\{\frac{\tilde{Y}_i}{1-\tilde{Y}_1-\tilde{Y}_2-\cdots-\tilde{Y}_{i-1}}\}_{i\geq 1}$ is distributed like i.i.d $\mathrm{beta}(1,\theta)$ if and only if $\{\tilde{Y}_i\}_{i\geq 1}$ is distributed like GEM with parameter θ if and only if $\{Y_i\}_{i\geq 1}$ is distributed $PD(\theta)$.

So, instead of working with $\left(\frac{\log p_1(N_{X,f})}{\log x}, \frac{\log p_2(N_{x,f})}{\log x}, \ldots\right)$, we work with the size-biased permutation of it. Turns out much more tractable.

Thank you for listening.