A Speculation of N. Katz Concerning
the Distribution of Arithmetic Quantities
(Notes by D. Wright of a lecture by N. Katz given at Princeton on Oct. 12, 1994)

1. Description of the Pattern

There are many important theorems in number theory asserting that a sequence of arithmetic quantities of some kind is uniformly distributed. N. Katz has discovered a remarkable pattern in the finer aspects of distribution of many of these sequences, which we will describe below.

Suppose for each \( n \geq 1 \) we have a set

\[ S_n = \{ 0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_{\nu_n} \leq 1 \} \]

of “angles” (normalized so that \( 2\pi = 1 \)). We shall assume that \( \lim_{n \to \infty} \nu_n = +\infty \). We say this set sequence is uniformly distributed if for any continuous \( \mathbb{C} \)-valued function \( f(\theta) \) on \( \mathbb{R}/\mathbb{Z} \) we have

\[ \lim_{n \to \infty} \frac{1}{\nu_n} \sum_{i=1}^{\nu_n} f(\theta_i) = \int_0^1 f(\theta) \, d\theta \]

Katz gave four examples of such sequences.

**Gauss sum angles:** The Gauss sums are

\[ g(\chi) = \sum_{x=1}^{p-1} \chi(p) \exp \left( 2\pi i \frac{x}{p} \right) \]

where \( p \) is an odd prime and \( \chi \) is a nontrivial multiplicative character modulo \( p \). It is well known that

\[ g(\chi) = \sqrt{p} \exp(2\pi i \theta_\chi) \]

for some angle \( \theta_\chi \) normalized as above. Thus, \( S_n \) consists of all the \( \theta_\chi \)'s for the \( n \)-th odd prime \( p \).

**Kloosterman sum angles:** The Kloosterman sums are

\[ \kappa(a) = \sum_{x=1}^{p-1} \exp \left( 2\pi i \frac{x + a\pi}{p} \right) \]

where \( a \in \mathbb{F}_p \) (the finite field of order \( p \)), \( a \neq 0 \), and \( \pi \) denotes the multiplicative inverse of \( x \) modulo \( p \). By the work of Hasse and Weil, it is known that

\[ \kappa(a) = 2\sqrt{p} \cos(\alpha_a) \]

for some angle \( \alpha_a \in [0, \pi] \). We would like to consider the set of \( \alpha_a \) for \( 1 \leq a \leq p-1 \), for the \( n \)-th prime \( p \). These are now known to be equidistributed with respect to the Sato-Tate measure thanks to work of Adolphson and Sperber and independently N. Katz. In this note, we “straighten out” these angles by the map:

\[ \theta_a = \frac{2\alpha_a - \sin(2\alpha_a)}{2\pi} \]

The \( \theta_a \)'s now form a uniformly distributed set sequence.
**Elliptic Curve Angles:** For an elliptic curve $E$ defined over $\mathbb{Q}$, let $N_E(p)$ be the number of points modulo $p$. Again by Hasse, we know that for the primes $p$ with good reduction

$$N_E(p) = 1 + p - a_p$$

where

$$a_p = 2\sqrt{p}\cos(\alpha_p)$$

with $\alpha_p \in [0, \pi]$. Conjecturally, these have the Sato-Tate distribution again. As in the Kloosterman case, we straighten the angles by the same formula to arrive at angles $\theta_p \in [0, 1]$. The $n$-th set then consists of $\theta_p$ for the first $n$ primes $p$ where $E$ has good reduction.

**Gaussian primes:** For any integer $n \geq 1$, let $S_n$ consist of $\theta_\eta = \text{Arg}(\eta)/2\pi$ for all primes $\eta \in \mathbb{Z}[i]$ with norm less than or equal to $n$, where $\text{Arg}$ denote the principal value of the argument of a complex number. Hecke showed by his theory of $L$-functions attached to Grössencharakteren that these angles are uniformly distributed.

We now suppose we have the angles in $S_n$ ordered in increasing fashion as described previously. We consider the normalized “spacings”

$$\delta_i = \nu_n(\theta_i - \theta_{i-1}), \quad 2 \leq i \leq \nu_n$$

and

$$\delta_1 = \nu_n(\theta_1 - \theta_{\nu_n} + 1)$$

These are a set of nonnegative real numbers with mean normalized to be 1. We define the cumulative distribution function to be

$$F_n(t) = \frac{1}{\nu_n} \text{Card}\{i \mid \delta_i \leq t\}$$

for all $t \geq 0$. Then $F_n$ is an increasing nonnegative step function such that $F_n(t) = 1$ for all $t$ sufficiently large.

The fantastic discovery of Katz is that in all cases described above we have empirically that

$$\lim_{n \to \infty} F_n(t) = 1 - e^{-t}$$

He produced many graphs of involved cases demonstrating this phenomenon.

This observed distribution would indicate a blurring of the normalized spacings. If the angles were exactly equally spaced the cumulative distribution function would be exactly

$$F(t) = \begin{cases} 
0, & \text{for } 0 \leq t < 1, \\
1, & \text{for } 1 \leq t
\end{cases}$$

No explanation, even conjectural, of the observed limit has apparently been offered by anyone.

**2. Some numerical computations**

The Gaussian prime angles are perhaps easiest to calculate in large quantities. Also, because of the vast amount of information provided by Hecke $L$-functions (they were invented by Hecke for an attack on the problem of proving that there are infinitely many primes of the form $n^2 + 1$, among other things), this might be the case with the strongest possibilities for attack. We used Henri Cohen’s PARI to tabulate some data.
First, we give a bound $n$ on the norms of the primes to be considered. For each prime $p \equiv 1 \pmod{4}$ with $p \leq n$, we calculated $0 < a < b$ such that $p = a^2 + b^2$. Then the Gaussian prime $b + ai$ has norm $p$ and argument between $0$ and $\pi/4$. By symmetry, we may restrict attention to Gaussian primes in that sector, as all of them are obtained through the variations $\pm a \pm bi$ and $\pm b \pm ai$, resulting merely in a multiplication of the numbers of spacings by 8. We then compute the associated angle

$$\theta_p = \arctan(a/b)$$

After this computation we have a large vector of these angles, which we sort in increasing order

$$0 < \theta_1 < \theta_2 < \cdots < \theta_l < \pi/4$$

(Note: all the $\theta$’s must be distinct by primality.) $l$ is the number of primes $p \equiv 1 \pmod{4}$ not greater than $n$.

We then compute the normalized spacings by

$$\delta_i = \frac{4l}{\pi} (\theta_i - \theta_{i-1})$$

for $1 \leq i \leq l$, where we interpret $\theta_0 = \theta_l - \pi/4$. Next these spacings are sorted into increasing order (and renumbered to reflect this sort). The cumulative distribution function is the step function

$$F(t) = \begin{cases} 
0, & \text{for } 0 \leq t < \delta_1, \\
\frac{i}{l}, & \text{for } \delta_i \leq t < \delta_{i+1}, \\
1, & \text{for } t \geq \delta_l
\end{cases}$$

In the document “Calculations of Gaussian primes,” we plot this function together with $1 - e^{-t}$ for $n = 100,000$ and $n = 500,000$. The agreement is plausible, but not totally convincing. The PARI programs are also listed in that document.

Concerning the primes $p \equiv 3 \pmod{4}$, there are only 65 of norm less than 500,000 (this requires $p^2 \leq 500,000$). Since there are 20,796 gaussian primes with norm less than 500,000 in that sector, including those primes adds only about 0.003 to the graph.

### 3. Quantifying the Agreement Between Prediction and Observation

Katz measured the quality of the match between theory and experiment by computing the Kolmogorov-Smirnov statistic as follows. Suppose we have computed the cumulative distribution function $F(x)$ on the basis of $N$ sample points. Supposed the expected distribution function is $G(x)$. Then the Kolmogorov-Smirnov statistic is the maximum deviation $|F(x) - G(x)|$ over all the computed values multiplied by $\sqrt{N}$. In the table below, we abbreviate the associated Kolmogorov-Smirnov statistic by KS and the maximum deviation by $\Delta_{\text{max}}$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n$</th>
<th># primes</th>
<th>KS</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,000,000</td>
<td>39,175</td>
<td>13.5318</td>
<td>0.0684</td>
</tr>
<tr>
<td>1,000,000</td>
<td>1,100,000</td>
<td>3642</td>
<td>1.3153</td>
<td>0.0218</td>
</tr>
<tr>
<td>5,000,000</td>
<td>6,000,000</td>
<td>32,138</td>
<td>4.0912</td>
<td>0.0228</td>
</tr>
</tbody>
</table>
4. **Origins**

The ideas and computations of Katz were motivated by recent work of Sarnak and Rudnick computing the $n$-tuple correlation function of the zeroes of the Riemann zeta function and more generally any automorphic $L$-function. This generalized the computation of the pair correlation of the Riemann zeta zeroes by Montgomery.