The Gaussian Zoo

John Renze, Stan Wagon, and Brian Wick

CONTENTS

We find all the maximal admissible connected sets of Gaussian primes: there are 52 of them. Our catalog corrects some errors in the literature. We also describe a totally automated procedure to determine the heuristic estimates for how often various patterns, in either the integers or Gaussian integers, occur in the primes. This heuristic requires a generalization of a classical formula of Mertens to the Gaussian integers, which we derive from a formula of Uchiyama regarding an Euler product that involves only primes congruent to 1 (mod 4).

1. INTRODUCTION

The twin prime conjecture states that the pattern $(n, n+2)$ occurs infinitely often in the prime numbers. On the other hand, $(n, n+2, n+4)$ cannot occur infinitely often because one of the entries will be divisible by 3. If we restrict patterns to the form $(n+a_1, n+a_2, \ldots, n+a_k)$, we can describe the pattern by the set $\{a_1, a_2, \ldots, a_k\}$; patterns in this context are called *constellations*. The question of which constellations occur infinitely often in the set of prime numbers has been well studied. Of course, this is unresolved even for $\{0, 2\}$, but it has been conjectured (the *prime k-tuples conjecture*) that if A has the property that for every prime p the number $r_p(A)$ of distinct entries in the mod-p reduction of A is less than p, then A occurs infinitely often in the primes. A set with this mod-p property for all primes is called *admissible*. Thus $\{0, 2\}$ is admissible, but $\{0, 2, 4\}$ is not because its reduction mod 3 is $\{0, 2, 1\}$. This paper extends the investigation of admissible sets to the Gaussian integers $\mathbb{Z}[i]$, a project started in [Jordan and Rabung 1976] and continued in [Vardi 1998].

For a Gaussian integer z, let $N(z)$ denote the norm of z: $N(a+bi) = a^2+b^2$. Recall that $a+bi$ is a *Gaussian prime* if a^2+b^2 is prime or one of a, b is 0 and the other is a prime congruent to 3 mod 4. Since Gaussian integers have unique factorization,

1058-6458/2001 $0.50 per page
Experimental Mathematics **10**:2, page 161

the definition of admissibility can be easily adapted. There are several ways to form a reduced residue system modulo a Gaussian prime p. We will use the following one [Ireland and Rosen 1982; Jordan and Potratz 1965]: If p is not a rational prime, we take the integers $\{0, 1, \dots, N(p)-1\}$ as the reduced system; for the rational Gaussian primes, we use $\{a+bi : 0 \le a, b \le p-1\}$. We will denote by $r_p(A)$ the cardinality of the mod-p Gaussian reduction of A, just as in the $\mathbb{Z}$ case; it will be clear from the context which version is meant.

Definition. A set A of Gaussian integers is *admissible* if $r_p(A) < N(p)$ for every Gaussian prime p. A prime p is a *blocking prime* of A if $r_p(A) = N(p)$.

So that we can easily visualize the patterns, we identify a Gaussian integer with a disk centered at it. Because we are primarily interested in the networks arising from Gaussian primes at distance $\sqrt{2}$ or less, we will use disks with radius $\sqrt{2}/2$. For the network arising from a larger distance, one uses a larger radius. We will use the term *animal* for a connected set of disks, in analogy with a term often used for connected sets in a lattice. Continuing the work of [Jordan and Rabung 1976; Vardi 1998], we will give a complete catalog of the admissible animals. Because, as shown by Jordan and Rabung, the largest has size 48, any admissible animal is contained in a maximal admissible animal.

Definition. A maximal admissible connected set of Gaussian integers is called a *lion.*

It follows that we need only consider the lions, because any admissible animal is contained in a maximal one. Up to symmetry, there are 52 lions.

It is natural to wonder how often one can expect various admissible patterns to occur in the Gaussian primes. There are well-known techniques to come up with heuristic estimates for such patterns in the rational primes, and we were able to extend these ideas to the Gaussian context. Perhaps more interesting, we have automated, in *Mathematica*, the production of such functions in both the rational and Gaussian cases. The formulas, which will be discussed in detail in Section 4, involve

$$\int_0^x \frac{1}{\log^k t}\,dt \qquad \text{or} \qquad \int_0^x \frac{t}{\log^k t}\,dt,$$

which can be expressed in terms of $\operatorname{li} x$, the logarithmic integral function. Then one can either express the formulas in a simple asymptotic form or, for more exact work, compute values of $\operatorname{li} x$ via the series $\operatorname{li} x = \gamma + \log\log x + \sum_{k=1}^{\infty} \log^k x/(k!\,k)$, where γ is Euler's constant [Riesel 1985, p. 55].

For example, we can ask our *Mathematica* program for the approximation to the number of twin primes less than x, by typing the command

```
ConstellationEstimate[{0,2}][x].
```

The result is

$$2\left(\prod_{p\ge 3}\frac{(p-2)\,p}{(p-1)^2}\right)\left(\operatorname{li} x - \frac{x}{\log x}\right),$$

or, in asymptotic form,

$$2\left(\prod_{p\ge 3}\frac{(p-2)\,p}{(p-1)^2}\right)\left(\frac{x}{\log^2 x}\right) + O\left(\frac{x}{\log^3 x}\right).$$

Immediate numerical evaluation of the product is possible, giving a coefficient of 1.32032 (with more digits if one wants). As another example, one can query the program for the function that approximates the number of prime patterns of the form $\{k, k+2, k+6, k+8\}$ with $k < x$; the result is

$$4.15118\left(\frac{\operatorname{li}(x)}{6} - \frac{x}{6\log x} - \frac{x}{6\log^2 x} - \frac{x}{3\log^3 x}\right) = 4.15118\,\frac{x}{\log^4 x} + O\left(\frac{x}{\log^5 x}\right).$$

The main benefits of using *Mathematica* for this project are the ease of use of Gaussian integer arithmetic, and the adaptive-precision capability, which eases the problem of getting reliable approximations to the infinite series that arise.

2. CAPTURING THE LIONS

The starting point of our lion hunt is [Vardi 1998], where it was shown, by working in the graph whose vertices are Gaussian integers relatively prime to 390 ($= 2\cdot 3\cdot 5\cdot 13$), how to compute 506 animals such that every admissible animal is contained in one of these. Let's be precise about the connection between such graphs and admissibility. Consider the graph made up of Gaussian integers relatively prime to some n, with edges connecting vertices at distance $\sqrt{2}$ or less. We claim that an admissible animal A must be congruent to a subset of a connected component of this graph. For if p is a Gaussian prime factor of n, then we know that when A is reduced modulo p, one of the residues is missing. This means that there is a Gaussian integer, m_p, such that 0 does not occur in the mod-p residues of $A+m_p$. Find such integers for each prime p and then use the Chinese Remainder Theorem to find a single integer m such that $A+m$ has no element divisible by any of the primes dividing n. It follows that each member of $A+m$ is coprime to n, as claimed.

One can reduce the set of 506 animals to 115 by using the eightfold symmetry in the Gaussian primes to eliminate duplicates. A further reduction is then possible by eliminating patterns that are congruent to subsets of other ones; this step leaves only 16 patterns. And 11 of these 16 turn out to be admissible (and hence are themselves lions). This leaves only five patterns whose subsets must be searched for lions. We call these five sets *protolions*; their sizes are 25, 37, 50, 51, 71 (Figure 1).

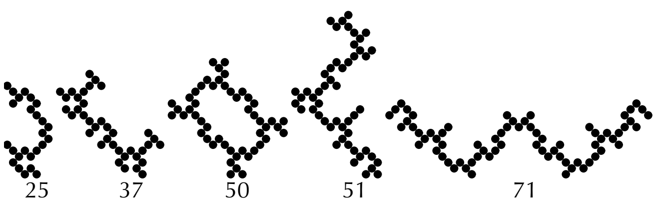

FIGURE 1. The five protolions.

Our approach to these protolions is as follows:

Step 1. Create two lists. The first list is called `Lions`; it starts with the 11 lions mentioned earlier, and eventually will contain all the lions. The second list is called `ToDo`, and initially it contains the 5 protolions mentioned earlier. At all times this list will contain animals, i.e., connected components.

Step 2. Repeat step 3 until the `ToDo` list is empty.

Step 3. Remove from the `ToDo` list any one of the connected components of maximal length (size). If it is admissible, then determine whether it is congruent to a subset of a lion in `Lions`. If it is, then discard it; otherwise add it to `Lions`. If, on the other hand, the set is not admissible then perform step 4.

Step 4. Determine a blocking prime p for the set. Repeat step 5 for each residue class modulo p and when finished, return to step 3.

Step 5. For a mod-p residue class, remove every element of the set with that residue. Find the connected components of the resulting set and place them on the `ToDo` list.

It is important to carry out step 3 efficiently. Given two animals, A and B, how does one determine whether A is congruent to a subset of B? Our routine is based on a technique used in digital image processing to find and describe the boundary of a connected region by its "chain code"; see [Gonzalez and Woods 1992], for example. A reason for using the chain code is that it is a sequence of digits that allows one to easily trace the boundary of a connected region or a boundary that is equivalent under symmetry. The digit k in the chain code is to be interpreted as the boundary turning by an angle of $45k$ degrees. For example, let $A = \{0,\ 1,\ 2,\ 3,\ 3+i,\ 2+2i,\ 1+i,\ i\}$ (Figure 2). The dots are the points of the set A, and the piecewise linear curve is the path taken to produce the chain code. The chain code for this region is 0, 0, 2, 1, 2, 7, 2, 2.

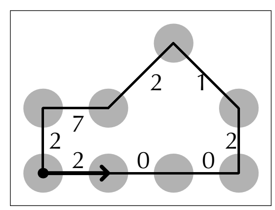

FIGURE 2. Animal with chain code 0, 0, 2, 1, 2, 7, 2, 2.

To reconstruct the animal from the chain code, an initial point and direction must be given. Suppose that the initial point is 0 and direction is parallel to the x-axis. Thus the second point will be 1. The chain code 0, 0, 2, 1, 2, 7, 2, 2 is to be interpreted as moving to the next Gaussian integer after turning 0, 0, 90, 45, 90, 315, 90, and 90 degrees, resp. Doing so, the algorithm will produce the set A. For a second example, if the initial point is 0 and direction is parallel to the y-axis, the chain code will generate the set $\{0, i, 2i, 3i, 3i-1, 2i-2, i-1, -1\}$. To obtain the chain code for the set that has been reflected about the x-axis, one replaces the digit k with $8-k$ unless $k = 0$ in which case the value of k doesn't change. Thus the chain code for the reflected set in our example is 0, 0, 6, 7, 6, 1, 6, 6. If the initial point is 0 and the initial direction is parallel to the x-axis, then the turns will be 0, 0, 270, 315, 270, 45, 270, and 270 degrees and the set produced will be $\{0, 1, 2, 3, 3-i, 2-2i, 1-i, -i\}$.

There was a point about which we had to be careful when coding the search routine. We had to verify that the chain code described the entire set, i.e., every point of the set is a boundary point. This is the case for all the animals found in step 3. Using chain codes, the algorithm to determine whether or not an animal A can be found in the animal B is easily implemented. The size of A must be less than or equal to the size of B. If so, then one checks for every member of the set B and for every initial diagonal direction, whether or not the chain code for A can step through elements of B.

Figure 4 shows the results of our search. There are 52 lions, the largest has size 48, there are seven of size 48, and the only ones that are not simply connected are those of size 4, 12, 47, and two of the 48s. The two lions of largest diameter are the ones of size 45 and 42. The seventh 48-lion (in the order of Figure 4) was missed in the computations of [Jordan and Rabung 1976] and [Vardi 1998].

We will use the term *diamond* to refer to the 4-lion and *castle* for the 12-lion. The order of the lions in Figure 4 is according to their rarity, as explained in Section 5.

To understand how the seven 48-lions arise, consider the 50-protolion and its two blocking primes, $2\pm5i$. Look at Figure 3, which shows the congruence classes modulo $2\pm5i$ (using the complete residue system $\{0, 1, \ldots, 28\}$). The gray disks correspond to entries that appear once only; the removal of any of these will fix the admissibility situation modulo the corresponding prime. Thus one can remove, say, the entry corresponding to 25 on the left and 21 on the right to get an admissible 48-set. There are, up to symmetry, six ways of doing this so as to leave a connected set. But the seventh 48-lion arises in an unusual way. Removal of the 13 from the left deletes an 11 on the right. There is only one other 11, so its removal will eliminate 11 entirely from the right-hand set, the result being that the two deletions lead to an admissible 48-set. Up to symmetry this type of double deletion leads to only one connected 48-set, and that is where the seventh 48-lion comes from.

Sections 3, 4, and 5 will address the question of how many diamonds or castles one can expect to find in a given region. Figures 5 and 6 give an idea of how numerous the diamonds are in the Gaussian primes. Note that diamonds that straddle the $y = x$ line are, in a sense, fake: the chances of their existence are greater than for the others because of the forced symmetry around the diagonal line. More precisely, if $a + (a-1)i$ and $a + 1 + a\,i$ are prime, then their reflections in the $y = x$ line are automatically prime, giving us a diamond for which the four primality events are not independent. We will discuss this in more detail in Section 5.

Since a brute force search for a castle might take an extremely long time, a filter was found and used to speed things up. After some experimentation,

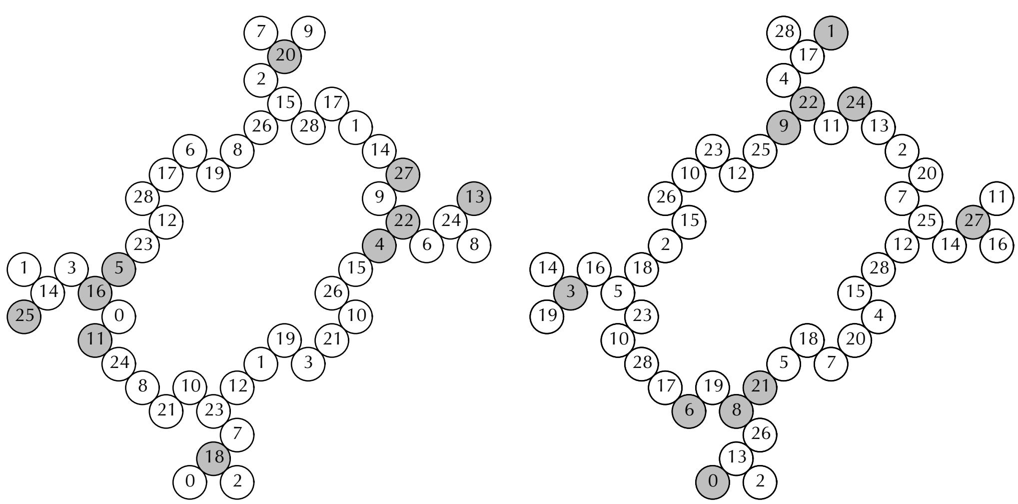

FIGURE 3. The reduction of the 50-protolion modulo the primes $2+5i$ (left) and $2-5i$ (right).

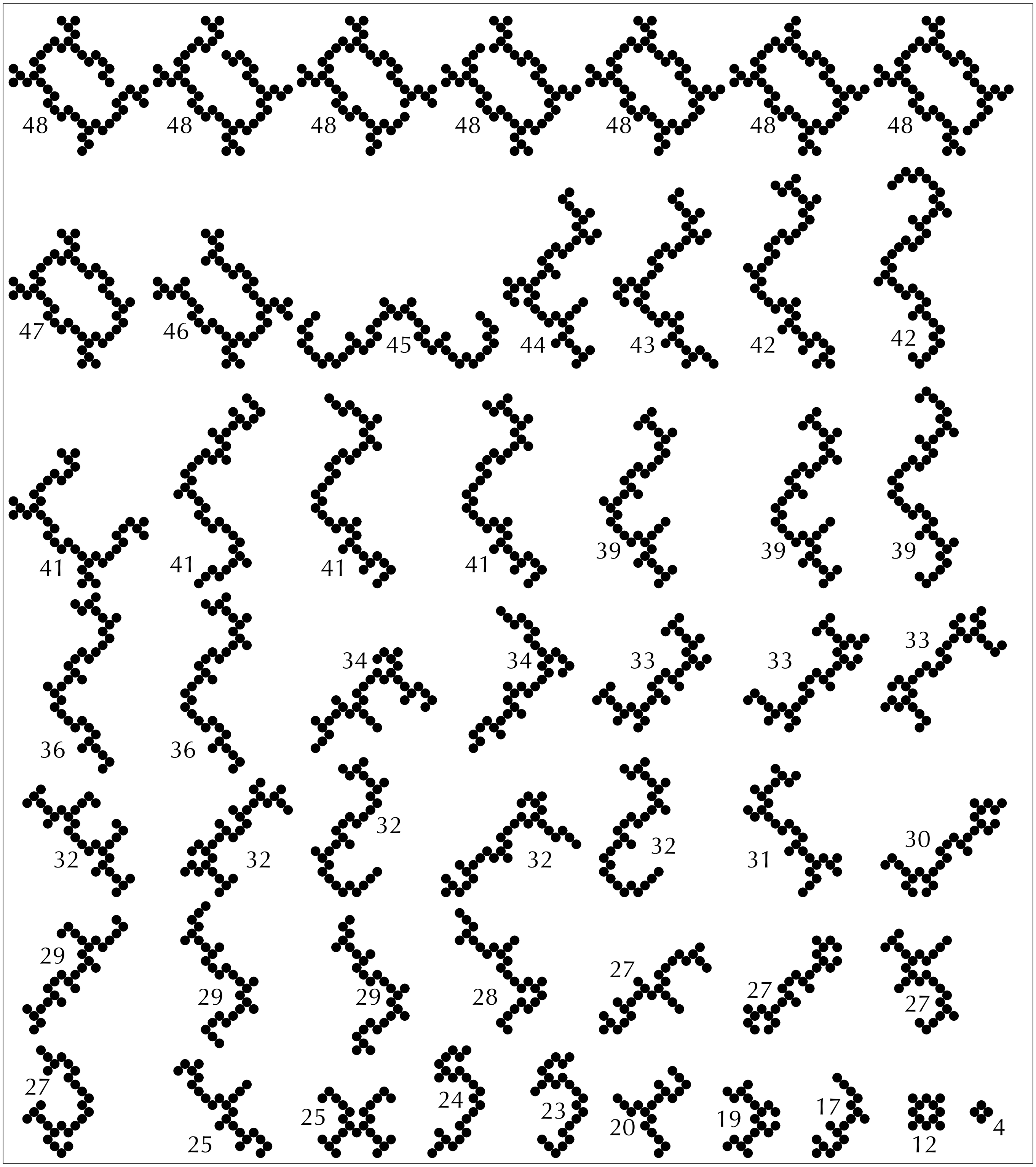

FIGURE 4. The complete catalog of 52 lions. All admissible animals are congruent to a subset of one of these.

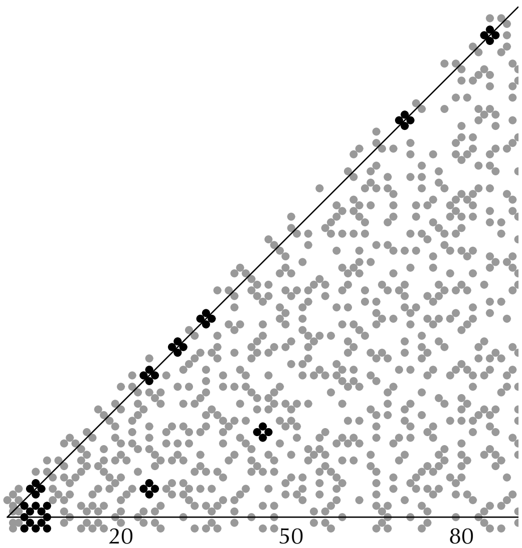

FIGURE 5. Two true diamonds and six fake diamonds, along with one castle that is fake because of the symmetry over the real axis. Some additional diamonds, both real and fake, can be found near the origin and are not specifically marked.

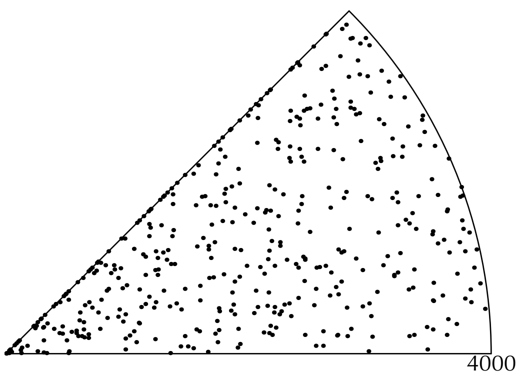

FIGURE 6. The centers of all the diamonds in the first octant of radius 4000; there are 380 true diamonds and 61 fake ones (centers on the 45° line).

we found that a square of size 6630×6630 had only 50 protocastles, where a *protocastle* is a connected set of 12 Gaussian integers relatively prime to 6630 that looks like a castle. Realizing that if a castle that is farther than distance 17 from the origin is reduced modulo 6630, then this reduction will be one of these protocastles, we were able to use this square to tile the plane and to reduce the search to checking only 50 protocastles in each tile. This search found a non-fake example of a castle centered at $7743840 + 4598295\,i$; see Figure 7. There is another castle centered at 5, but its existence relies on

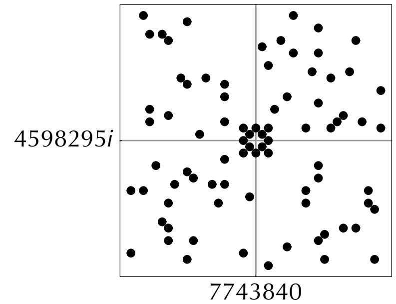

FIGURE 7. The first real castle in the Gaussian primes.

symmetry about the x-axis. (Note: The tile containing this fake castle had to be independently searched because one of its primes is a factor of 6630.)

To generalize our work, one should consider admissible sets that are connected when edges are determined by distance 2 or less. Gethner and Stark [1997] showed that there are only finitely many admissible sets in this case. We verified their results by considering, as they did, the distance-2 graph whose vertices are the Gaussian integers that are coprime to n, where n is 7113990 ($= 2 \cdot 3 \cdot 5 \cdot 13 \cdot 17 \cdot 29 \cdot 37$). We took a slightly different approach, but confirmed their main computational result: Every connected set in the graph that crosses the real axis is finite and has imaginary parts bounded by ± 357. Periodicity and symmetry then imply that any connected set in the graph is finite.

We then tried to find lions in the distance-2 case. The top-down approach we used for distance $\sqrt{2}$ worked because of the modest size of the protolions (Figure 1). The corresponding animals are too large in the distance 2 case, so we adopted a slower, but conceptually simpler, bottom-up approach. Start with a single point. Add potential neighbors until an animal is found that is admissible but has no admissibe extension to an animal with one more element. That will be a lion. This also provided a check on our distance-$\sqrt{2}$ work, since it yielded the identical set of lions that the top-down method found. Computation then showed that there are no lions for distance 2 of size 15 or smaller. However, the patterns that arose led us to a conjecture about a 16-lion (Figure 8), and it was then easy to check that it is in fact a lion, perhaps the unique smallest one in the distance-2 case.

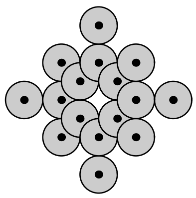

FIGURE 8. A size-16 lion in the distance-2 case: the castle with four points added.

3. ESTIMATING THE NUMBER OF PRIME CONSTELLATIONS

Once a constellation is known to be admissible, a natural question to ask is how often it occurs. Questions of this type are notoriously difficult, but we can use ideas from the rational case to develop a heuristic estimate for the frequency of a given Gaussian constellation. In Section 5, we will provide numerical evidence for our estimates.

We first review the classical approach, carried out by, among others, Hardy and Littlewood [1923] (see also [Hardy and Wright 1960, §22.20]), to estimate the number of times a prime constellation occurs in a given interval. Let $A = \{a_1, \ldots, a_n\}$ be an admissible constellation and let $\mathcal{P}$ denote the rational primes. We seek a heuristic formula for the probability that an integer x starts a prime A-pattern—i.e., that $x+A \subseteq \mathcal{P}$—which we can then integrate to get the expected number of prime occurrences of A in an interval. For connections between such asymptotic estimates and the broader theory of percolation in lattices, see [Vardi 1999].

A precise treatment calls upon the important theorem of Mertens [Hardy and Wright 1960, §22.8; Tenenbaum and Mendès France 2000], which states that the Euler product $\prod_{p\le x}\left(1-1/p\right)$ is asymptotic to $c/\log x$, where $c = e^{-\gamma} \approx 0.56$. Some explanation of the constant c is in order.

A naive approach to prime counting uses a simple random model to simulate the sieve of Eratosthenes, arguing that half of the numbers under x are divisible by 2, one third of the remainder are divisible by 3, and so on. This leads to the estimate $\prod_{p\le\sqrt{x}}\left(1-1/p\right)$ for the prime density near x. But this formula is wrong: by Mertens's theorem, this product is asymptotic to $2c/\log x$, or $1.12\ldots/\log x$, contradicting the Prime Number Theorem. In short, the sieve of Eratosthenes is about 11% more efficient than randomness would predict. The point is that divisibility by different primes is not as independent as one might expect. An observation in [Furry 1942] gives some insight: "the last and largest of the trial divisors finds roughly twice as large a proportion of victims among the survivors of previous trials as it would in a virgin population." His statement can be justified if we assume that the number of primes under x is well approximated by $kx/\log x$, for some constant k. Then, if x is not a square and p is the largest prime below $\sqrt{x}$, the potential victims between p^2 and x are either primes or numbers of the form $p\cdot q$, with q prime and $q \ge p$. The latter are the final victims of the sieve. Therefore the deletion ratio for p-divisibility is

$$\frac{\pi(x/p)-\pi(p)}{\pi(x)-\pi(p^2)+\pi(x/p)-\pi(p)}.$$

If we approximate this by using $k/\log x$ to estimate for the prime density near both x and p^2 and $k/\log\sqrt{x}$ for the prime density near both p and x/p, we get

$$\frac{\dfrac{k}{\log\sqrt{x}}\left(\dfrac{x}{p}-p\right)}{\dfrac{k}{\log x}(x-p^2)+\dfrac{k}{\log\sqrt{x}}\left(\dfrac{x}{p}-p\right)}.$$

This simplifies to $2/(p+2)$, essentially twice the $1/p$ predicted by the random model. For a more modern treatment of the fine points of the sieve of Eratosthenes, see the discussion of the Buchstab function in [Friedlander et al. 1991].

Now let's see how the modified random model helps us get an estimate of the number of prime constellations. We use p to denote a prime and Prob to denote probability; the use of "$\approx$" means "is heuristically asymptotic to", under the random model of divisibility:

$$\begin{aligned}
&\mathrm{Prob}[x+A\subseteq\mathcal{P}]\\
&\quad=\mathrm{Prob}[x+a_i \text{ is prime for } 1\le i\le n]\\
&\quad\approx\mathrm{Prob}[x+a_i \text{ is not divisible by } a \text{ prime under } x^c\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{for } \ 1\le i\le n]\\
&\quad=\mathrm{Prob}[x\not\equiv -a_i \pmod{p} \text{ for } 1\le i\le n \text{ and } p<x^c].
\end{aligned}$$

There are p possible mod-p residues of x and $r_p(A)$ of them are disallowed. Therefore the preceding probability can be estimated as follows, where the $\sim$ step uses Mertens's theorem to multiply by $1/\log x$

and also by the reciprocal of the product it is asymptotic to, both to the n-th power. The probability that $x \not\equiv -a_i \pmod p$ for $1 \le i \le n$ and $p \le x^c$ is

$$\begin{aligned}
&\approx \prod_{p \le x^c} \frac{p - r_p(A)}{p} \\
&\sim \frac{1}{\log^n x} \prod_{p \le x^c} \Big(\frac{p}{p-1}\Big)^n \prod_{p \le x^c} \frac{p - r_p(A)}{p} \\
&= \frac{1}{\log^n x} \prod_{p \le x^c} \frac{p^{n-1}(p - r_p(A))}{(p-1)^n} \\
&= \frac{1}{\log^n x} \prod_{p=2}^{\infty} \frac{p^{n-1}(p - r_p(A))}{(p-1)^n} = \frac{C_A}{\log^n x}. \qquad (3\text{–}1)
\end{aligned}$$

In the next-to-last step the product to x^c is extended to infinity, because the infinite product converges (this follows from work in Section 4) and so the tail is asymptotic to 1. This yields the constant C_A, called the *Hardy–Littlewood constant* for A. Now, switching to t as the variable of integration, we can use the integral $C_A \int_0^x 1/\log^n t\, dt$ as an estimate of the number of prime occurrences of A in $[1, x]$. Because $\int_0^x 1/\log^n t\, dt$ asymptotic to $x/\log^n x$ (integration by parts), this leads to the Hardy–Littlewood conjecture.

The Prime k-Tuples Conjecture. Any admissible set A of n integers occurs infinitely often in the primes. If C_A denotes the Hardy–Littlewood constant for A, the number of occurrences below x is asymptotically equal to

$$C_A \, \frac{x}{\log^n x}.$$

To use the random model to estimate the number of constellations in the Gaussian primes, we must first generalize Mertens's theorem. Letting p_i denote a prime congruent to $i \pmod 4$, we define the *Gaussian Euler product for r* as

$$\prod_{N(p) \le r^2} \Big(1 - \frac{1}{N(p)}\Big),$$

which equals

$$(1 - \tfrac{1}{2}) \prod_{p_1 \le r^2} \Big(1 - \frac{1}{p_1}\Big)^2 \prod_{p_3 \le r} \Big(1 - \frac{1}{p_3^2}\Big).$$

To study Mertens's formula in $\mathbb{Z}[i]$, one asks if the Gaussian Euler product for r^c is asymptotic to the Gaussian prime density, which is $2/(\pi \log r)$, as proved in [Vardi 1998]. Here is the generalization.

Theorem (Mertens's Theorem for Gaussian Integers). *The product of $1 - 1/N(p)$ over Gaussian primes in the disk of radius r^c is asymptotic to the Gaussian prime density at radius r. That is,*

$$\prod_{N(p) \le r^{2c}} \Big(1 - \frac{1}{N(p)}\Big)$$

is asymptotic to $2/(\pi \log r)$.

Proof. The key is a formula from [Uchiyama 1971] (see also the proof of [Vardi 1998, Proposition 2.1]):

$$\prod_{p_1 \le x} \Big(1 - \frac{1}{p_1}\Big) = \sqrt{\frac{c\,\pi\, d_1}{\log x}} + O\Big(\frac{1}{\log^{3/2}(x)}\Big),$$

where $d_1 := \prod_{p_1=5}^{\infty} (1 - 1/p_1^2)$ and $d_3 := \prod_{p_3=3}^{\infty} (1 - 1/p_3^2)$. Note that

$$(1 - \tfrac{1}{4}) d_1 d_3 = \prod_{p=2}^{\infty} \Big(1 - \frac{1}{p^2}\Big) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2},$$

so $d_1 d_3 = 8/\pi^2$. To quickly compute the numerical value of d_3 (it is $0.8561\ldots$), see [Vardi 1998].

Using Uchiyama's formula, we get

$$\begin{aligned}
\prod_{N(p) \le r^{2c}} \Big(1 - \frac{1}{N(p)}\Big) &= \frac{1}{2} \prod_{p_1 \le r^{2c}} \Big(1 - \frac{1}{p_1}\Big)^2 \prod_{p_3 \le r^c} \Big(1 - \frac{1}{p_3^2}\Big) \\
&\sim \frac{1}{2} \, \frac{c\,\pi\, d_1}{2c \log r} d_3 = \frac{\pi d_1 d_3}{4 \log r} \\
&= \frac{2}{\pi \log r}. \qquad \square
\end{aligned}$$

With this tool in hand, the development of a heuristic probability that an admissible Gaussian constellation A of size n occurs in the Gaussian primes starting at a Gaussian integer z is identical to the rational case, keeping in mind that there are $N(p)$ residues for any Gaussian prime p. Using r for $|z|$ and arguing exactly as in (3–1), the probability that $A + z$ is contained in the Gaussian primes is

$$\Big(\frac{2}{\pi \log r}\Big)^n \prod_{N(p) \le r^{2c}} \Big(\frac{N(p) - r_p(A)}{N(p)}\Big) \Big(\frac{N(p)}{N(p)-1}\Big)^n. \qquad (3\text{–}2)$$

And so the number of prime occurrences of A in the first octant and at radius R or smaller is estimated by

$$C_A\left(\frac{2}{\pi}\right)^n \frac{\pi}{4}\int_0^R \frac{r}{\log^n r}dr,$$

where C_A is the product in (3–2) (extended to an infinite product) and the standard polar-form factor r occurs in the integrand because the it arises from a double integral over the sector $0 \le \theta \le \pi/4$ and $0 \le r \le R$. This allows us to state the analog of the prime k-tuples conjecture for the Gaussian integers. Note that the definite integral just given is asymptotic to $R^2/(2\log^n R)$.

The Prime k-Tuples Conjecture in $\mathbb{Z}[i]$. Any admissible n-element set A of Gaussian integers occurs infinitely often in the Gaussian primes; the number of occurrences in a disk of radius R and within the first octant is asymptotic to

$$\frac{C_A\, 2^{n-3}\, R^2}{\pi^{n-1}\log^n R},$$

where C_A is the Gaussian Hardy–Littlewood constant for A.

4. COMPUTING THE HARDY–LITTLEWOOD CONSTANTS

The computation of numerical approximations to the Hardy–Littlewood constants is similar in the rational and Gaussian cases. We first review the rational case. Given A, let p_0 be the first prime greater than any difference of two members of $A = \{a_1, \dots, a_n\}$; therefore $r_p(A) = n$ whenever $p \ge p_0$. Then C_A from (3–1) is

$$\prod_{p<p_0} \frac{p^{n-1}(p-r_p(A))}{(p-1)^n}\prod_{p\ge p_0}\frac{p^{n-1}(p-n)}{(p-1)^n}.$$

The finite product is easy to get by a direct computation of the critical prime p_0 and each $r_p(A)$. And there are standard methods for getting numerical approximations to the infinite product. One uses logarithms as follows, where $\zeta_{\text{prime}}(s)$ is the prime zeta function, $\sum p^{-s}$; it is not hard to see that

$$\zeta_{\text{prime}}(s) = \sum_{k=1}^{\infty}\frac{\mu(k)}{k}\log\zeta(k\,s),$$

where μ is the Möbius function and ζ is the Riemann zeta function (see [Riesel 1985] for an outline and [Vardi 1991] for a detailed description that shows how variations of this technique can be used in diverse Euler product computations). Then

$$\begin{aligned}
\log&\left(\prod_{p=p_0}^{\infty}\frac{p^{n-1}(p-n)}{(p-1)^n}\right)\\
&=\sum_{p\ge p_0}\log\Bigl(1-\frac{n}{p}\Bigr)-n\log\Bigl(1-\frac{1}{p}\Bigr)\\
&=\sum_{p\ge p_0}\Bigl(-\frac{n}{p}-\frac{n^2}{2p^2}-\frac{n^3}{3p^3}-\cdots\\
&\qquad\qquad+\frac{n}{p}+\frac{n}{2p^2}+\frac{n}{3p^3}+\cdots\Bigr)\\
&=\sum_{p\ge p_0}\sum_{j=2}^{\infty}\frac{n-n^j}{j}\frac{1}{p^j}\\
&=-\sum_{j=2}^{\infty}\frac{n^j-n}{j}\Bigl(\zeta_{\text{prime}}(j)-\sum_{p<p_0}\frac{1}{p^j}\Bigr).
\end{aligned}$$

This final series is convergent, and the tail is easy to bound, because $p_0 \ge n$ and $\sum_{p<p_0} p^{-j}$ is bounded by (use an integral) $1/\bigl(n^{j-1}(j-1)\bigr)$; this justifies our earlier statement about the convergence of the infinite products. So we may now conclude that

$$\begin{aligned}
C_A &= \left(\prod_{p\le p_0}\frac{p^{n-1}(p-r_p(A))}{(p-1)^n}\right)\\
&\times\exp\left(-\sum_{j=2}^{\infty}\frac{n^j-n}{j}\Bigl(\zeta_{\text{prime}}(j)-\sum_{p<p_0}\frac{1}{p^j}\Bigr)\right). \qquad (4\text{–}1)
\end{aligned}$$

For the Gaussian case we first define n_0 be the least integer such that if $N(p) \ge n_0$ the Gaussian prime p does not divide the difference between any two members of A. Then

$$\begin{aligned}
C_A &= \prod_{N(p)<n_0}\left(\frac{N(p)-r_p(A)}{N(p)}\right)\left(\frac{N(p)}{N(p)-1}\right)^n\\
&\times\prod_{N(p)\ge n_0}\left(\frac{N(p)-n}{N(p)}\right)\left(\frac{N(p)}{N(p)-1}\right)^n. \qquad (4\text{–}2)
\end{aligned}$$

To approximate the infinite product, we will need Gaussian versions of the zeta and prime zeta functions. For the first, set $\zeta_{\text{Gauss}}(s) := \sum_{n\in\mathbb{Z}[i]} N(n)^{-s}$. Computation is easy because

$$\zeta_{\text{Gauss}}(s) = \frac{1}{4^s}\zeta(s)\bigl(\zeta(s,\tfrac14)-\zeta(s,\tfrac34)\bigr),$$

where $\zeta(s, a)$ is the Hurwitz zeta function (or generalized zeta function) $\sum_{k=1}^{\infty}(k+a)^{-s}$. This can be easily proved using the infinite product form of ζ, as follows, where the last step comes from the usual geometric series interpretation of the infinite products:

$$\begin{aligned}
&\zeta_{\mathrm{Gauss}}(s)\\
&\quad=\sum_{n\in\mathbb{Z}[i]}\frac{1}{N(n)^s}=\prod_{p}\frac{1}{1-1/N(p)^{-s}}\\
&\quad=\frac{1}{1-2^{-s}}\prod_{p_3}\frac{1}{1-p_3^{-2s}}\Bigl(\prod_{p_1}\frac{1}{1-p_1^{-s}}\Bigr)^2\\
&\quad=\frac{1}{1-2^{-s}}\prod_{p_3}\frac{1}{1-p_3^{-s}}\prod_{p_3}\frac{1}{1+p_3^{-s}}\Bigl(\prod_{p_1}\frac{1}{1-p_1^{-s}}\Bigr)^2\\
&\quad=\prod_{\substack{\text{rational}\\ \text{primes } q}}\frac{1}{1-q^{-s}}\prod_{p_3}\frac{1}{1+p_3^{-s}}\prod_{p_1}\frac{1}{1-p_1^{-s}}\\
&\quad=4^{-s}\zeta(s)\bigl(\zeta(s,\tfrac14)-\zeta(s,\tfrac34)\bigr).
\end{aligned}$$

We also need the prime version, defined by

$$\zeta_{\mathrm{Gp}}(s)=\sum_{p\in\mathbb{Z}[i]}N(p)^{-s}.$$

By Möbius inversion as in the rational case, this equals

$$\sum_{k=1}^{\infty}\frac{\mu(k)}{k}\log\zeta_{\mathrm{Gauss}}(ks).$$

And just as in the rational case, the infinite product in (4–2) reduces to the sum

$$\begin{aligned}
&\exp\Bigl(\sum_{j=2}^{\infty}\frac{n^j-n}{j}\sum_{N(p)\ge n_0}\frac{1}{p^j}\Bigr)\\
&\qquad=\exp\Bigl(\sum_{j=2}^{\infty}\frac{n^j-n}{j}\Bigl(\zeta_{\mathrm{Gp}}(j)-\sum_{N(p)<n_0}\frac{1}{N(p)^j}\Bigr)\Bigr).
\end{aligned}$$

The methods of evaluating the infinite sums in (4–1) and (4–2) are similar. The infinite sum in (4–1) can be computed to any precision by using the Möbius function to compute ζ_{prime}. *Mathematica*'s adaptive precision — its algorithm that increases the working precision beyond d as necessary to get d significant digits — is helpful to handle the subtractive cancellation that occurs in the computation of ζ_{prime}. The adaptive precision algorithm uses some heuristics to estimate precision, and so is not foolproof. But the functions that arise here are simple (logarithms, multiplication, division, power) and one can check the results by using a high fixed precision; such checking gives us confidence that the method is accurate in the present problem. Of course, one must also do some error analysis to see where to cut off the infinite sums that occur, both in the sum as j goes from 2 to infinity and the sum involving the Möbius function. But this is quite routine, using only very elementary techniques to bound the size of the tail in each case.

Because the Riemann ζ function is easy to evaluate [Edwards 1974, Chapter 6], the Gaussian versions of ζ can be evaluated without difficulty. As in the rational case, the work needed to determine, simplify, and numerically evaluate the estimate of Gaussian prime constellations can be automated.

5. COMPUTATIONAL RESULTS

Here are sample outputs from our program. First, the symbolic form of the estimated number of twin primes below x:

$$2\Bigl(\prod_{p\ge 3}\frac{(p-2)p}{(p-1)^2}\Bigr)\Bigl(\operatorname{li}x-\frac{x}{\log x}\Bigr).$$

A numerical approximation:

$$1.32032\Bigl(\operatorname{li}x-\frac{x}{\log x}\Bigr).$$

The asymptotic form:

$$\frac{1.32032\,x}{\log^2 x}+O\Bigl(\frac{x}{\log^3 x}\Bigr).$$

The program's 14-digit approximation to the twin prime estimate for 10^{14} is 135780264892.06. (The actual number of twin primes is 135780321665.)

The program can be asked to return the answer in terms of unevaluated integrals, as in

$$2\Bigl(\prod_{p\ge 3}\frac{(p-2)p}{(p-1)^2}\Bigr)\int_0^x\frac{1}{\log^2 t}\,dt.$$

A famous admissible constellation [Riesel 1985, p. 78] is the collection of 15 primes starting at 11, namely {11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67}. No example of this full pattern has yet been found in the primes. Here is the estimate for the number of occurrences of the pattern below x:

$$\frac{187824\,x}{\log^{15}x}+O\Big(\frac{x}{\log^{16}x}\Big).$$

An examination of the more explicit form shows that the estimate equals 1 at $1.7\cdot 10^{19}$.

We now turn to the Gaussian integers. Here are four examples of the output of our program.

The diamond estimate in terms of an integral:

$$\frac{2^5}{\pi^3}\Bigg(\prod_{N(p)\geq 5}\frac{(N(p)-4)(N(p))^3}{(N(p)-1)^4}\Bigg)\int_0^R\frac{r}{\log^4 r}dr.$$

With the integral evaluated and the coefficient approximated numerically (the two steps can be carried out separately):

$$0.177975\left(\tfrac{4}{3}\,\mathrm{li}\,R^3-\frac{2\,R^2}{3\log R}-\frac{R^2}{3\log^2 R}-\frac{R^2}{3\log^3 R}\right).$$

The asymptotic form:

$$\frac{0.059325\,R^2}{\log^4 R}+O\Big(\frac{R^2}{\log^5 R}\Big).$$

An estimate of the number of diamonds within distance 200000 of the origin: 193922.

We have counted the number of diamonds in the first octant and within distance 200000 of the origin and the comparison to the predicted value is good. Before presenting the data we note a complication. The heuristic formulas are based on only the diamonds that are contained entirely in the first octant. But there are (presumably) an infinite number of diamonds that straddle the $y=x$ line (such as $1+2i$, $3+2i$, $2+i$, $2+3i$) and must be counted. We call them *fake* diamonds. The condition for the occurrence of a fake diamond on the $y=x$ line is for the two values $a+(a-1)i$ and $(a+1)+a\,i$ to be prime. Then symmetry guarantees that the other two points in the diamond are prime. No diamonds straddle the real axis, because there are no Gaussian twins in the integers. The extra complexity that the fakes cause in the diamond case is not great, and we can easily estimate the expected number of all diamonds, both true and fake. But if one wanted similar estimates for larger admissible sets, the symmetry issues would become more complicated because of the different ways a pattern can straddle the line of symmetry. For the case at hand, the number of fake diamonds in the first quadrant and within distance R of the origin is estimated by the following formula, which uses the techniques discussed earlier to estimate the number of Gaussian prime occurrences of the pattern $\{0, 1+i\}$ in the set $\{(a+1)+a\,i : 0\leq a\leq R\}$:

$$2\Bigg(\prod_{\substack{p\text{ a Gaussian}\\ \text{prime}\neq 1+i}}\frac{(N(p)-2)\,N(p)}{(N(p)-1)^2}\Bigg)\Bigg(\int_0^R\Big(\frac{2}{\pi\log r}\Big)^2 dr\Bigg).$$

This is $O(R/\log^2 R)$, which is asymptotically zero when divided by the expected number of true diamonds.

When R is 200000, the real and fake diamond estimates are 193922.35 and 1121.69, respectively, for a total of 195044. In fact, there are 193628 real diamonds and 1145 fake diamonds, for a total of 194773. Figure 9 shows the ratios of the actual

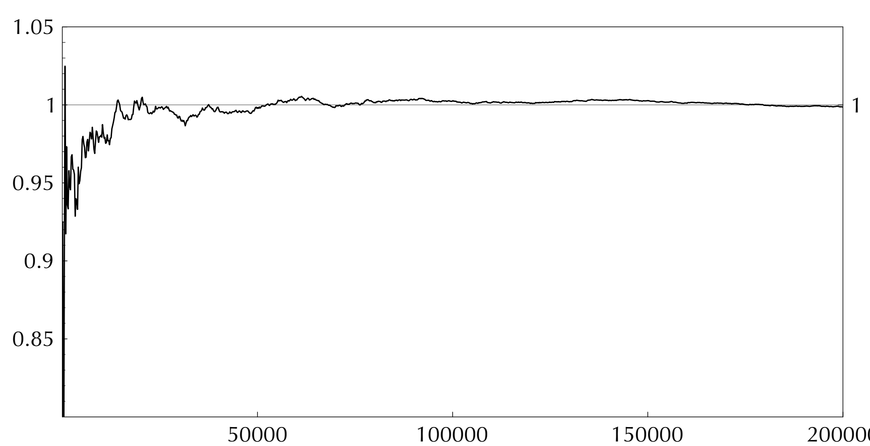

FIGURE 9. The ratio of the total number of diamonds at radius R or less to the predicted number. The ratio appears to converge to 1, as expected.

number of diamonds to the predicted number; the convergence to 1 is evident.

Our program gives the symbolic form of the prediction function for the next larger lion, the castle:

$$\frac{5^{24}\,13^{22}\,17^{24}\,29^{22}}{\pi^{11}\,2^{255}\,3^2\,7^{24}}\left(\prod_{N(p)\geq 37}\frac{(N(p)-12)\,(N(p))^{11}}{(N(p)-1)^{12}}\right)$$
$$\times\left(\frac{8\,\mathrm{li}\,R^2}{155925}-\frac{4\,R^2}{155925\,\log R}-\frac{2\,R^2}{155925\,\log^2 R}\right.$$
$$-\frac{2\,R^2}{155925\,\log^3 R}-\frac{R^2}{51975\,\log^4 R}-\frac{2\,R^2}{51975\,\log^5 R}$$
$$-\frac{R^2}{10395\,\log^6 R}-\frac{R^2}{3465\,\log^7 R}-\frac{R^2}{990\,\log^8 R}$$
$$\left.-\frac{2\,R^2}{495\,\log^9 R}-\frac{R^2}{55\,\log^{10} R}-\frac{R^2}{11\,\log^{11} R}\right).$$

The asymptotic form is

$$\frac{0.09528\,R^2}{\log^{12} R}+O\left(\frac{R^2}{\log^{13} R}\right).$$

Straightforward root finding shows the radius R at which the expected number of castles is 1 to be $1.04\cdot 10^8$. This is reasonably consistent with our discovery (Figure 7) of the first true castle at distance $9\cdot 10^6$ from the origin.

Finally, we can look at the seven 48-lions and see which one is the rarest. Note that each of the 48 lions is, asymptotically, rarer than any smaller lion. The following list gives the Hardy–Littlewood constants for the order in which the 48-lions appear in Figure 3; the first one is the rarest: $6.32\cdot 10^{21}$, $6.84\cdot 10^{21}$, $6.89\cdot 10^{21}$, $6.89\ \cdot 10^{21}$, $6.9\cdot 10^{21}$, $6.9\cdot 10^{21}$, $6.91\cdot 10^{21}$.

ACKNOWLEDGEMENTS

We are grateful to Ilan Vardi for helpful pointers on the direction of this work, to a referee and editor for many helpful comments, and to David Bressoud for directing us to the Uchiyama paper.

ELECTRONIC AVAILABILITY

The *Mathematica* code used for computing Hardy–Littlewood estimates for admissible sets is available at http://www.expmath.org/extra/10.2/rww.

REFERENCES

[Edwards 1974] H. M. Edwards, *Riemann's zeta function*, Pure and Applied Mathematics **58**, Academic Press, New York and London, 1974.

[Friedlander et al. 1991] J. Friedlander, A. Granville, A. Hildebrand, and H. Maier, "Oscillation theorems for primes in arithmetic progressions and for sifting functions", *J. Amer. Math. Soc.* **4**:1 (1991), 25–86.

[Furry 1942] W. H. Furry, "Number of primes and probability considerations", *Nature* **150** (1942), 120–121.

[Gethner and Stark 1997] E. Gethner and H. M. Stark, "Periodic Gaussian moats", *Experiment. Math.* **6**:4 (1997), 289–292.

[Hardman and Jordan 1967] N. R. Hardman and J. H. Jordan, "A minimum problem connected with complete residue systems in the Gaussian integers", *Amer. Math. Monthly* **74** (1967), 559–561.

[Hardy and Littlewood 1922] G. H. Hardy and J. E. Littlewood, "Some problems of 'Partitio Numerorum' III: On the expression of a number as a sum of primes", *Acta Math.* **44** (1922), 1–70. Reprinted as pp. 561–630 of *Collected Papers of G. H. Hardy*, vol. 1, Oxford, Clarendon Press.

[Hardy and Wright 1960] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Fourth ed., Oxford University Press, New York, 1960. Fifth ed., 1979.

[Ireland and Rosen 1982] K. F. Ireland and M. I. Rosen, *A classical introduction to modern number theory*, GTM **84**, Springer, New York, 1982. Second ed., 1990.

[Jordan and Potratz 1965] J. R. Jordan and C. J. Potratz, "Complete residue systems in the Gaussian integers", *Math. Mag.* **38** (1965), 1–12.

[Jordan and Rabung 1976] J. H. Jordan and J. R. Rabung, "Local distribution of Gaussian primes", *J. Number Theory* **8**:1 (1976), 43–51.

[Riesel 1985] H. Riesel, *Prime numbers and computer methods for factorization*, Birkhäuser, Boston, 1985. Second ed., 1994.

[Tenenbaum and Mendès France 2000] G. Tenenbaum and M. Mendès France, *The prime numbers and their distribution*, Amer. Math. Soc., Providence, RI, 2000.

[Uchiyama 1971] S. Uchiyama, "On some products involving primes", *Proc. Amer. Math. Soc.* **28** (1971), 629–630.

[Vardi 1991] I. Vardi, *Computational recreations in Mathematica*, Addison-Wesley, Redwood City, CA, 1991.

[Vardi 1998] I. Vardi, “Prime percolation”, *Experiment. Math.* **7**:3 (1998), 275–289.

[Vardi 1999] I. Vardi, “Deterministic percolation”, *Comm. Math. Phys.* **207**:1 (1999), 43–66.

John Renze, Department of Mathematics, Northwestern University, Evanston, IL 60208, United States (jrenze@math.nwu.edu)

Stan Wagon, Department of Mathematics and Computer Science, Macalester College, St. Paul, MN 55105, United States (wagon@macalester.edu)

Brian Wick, Department of Mathematical Sciences, University of Alaska Anchorage, 3211 Providence Dr., Anchorage AK 99508 United States (afbdw@uaa.alaska.edu)

Received June 24, 1999; accepted in revised form June 20, 2000