Smooth Values of Quadratic Polynomials

J. B. Conrey & M. A. Holmstrom

EXPERIMENTAL MATHEMATICS
2021, VOL. 30, NO. 4, 447–452
https://doi.org/10.1080/10586458.2018.1559775

Smooth Values of Quadratic Polynomials

J. B. Conrey[a,b] and M. A. Holmstrom[c]

[a]American Institute of Mathematics, San Jose, CA, USA; [b]University of Bristol, Bristol, UK; [c]Stanford University, Stanford, CA, USA

ABSTRACT
Let $\mathcal{Q}_a(z)$ be the set of z-smooth numbers of the form q^2+a. It is not obvious, but this is a finite set. The cardinality can be quite large; for example, $|\mathcal{Q}_1(1900)| \geq 646890$. We have a remarkably simple and fast algorithm that for any a and any z yields a subset $Q_a(z) \subset \mathcal{Q}_a(z)$ which we believe contains all but a tiny fraction of the elements of $\mathcal{Q}_a(z)$, i.e. $|\mathcal{Q}_a(z)| = (1+o(1))|Q_a(z)|$. We have used this algorithm to compute $Q_a(500)$ for all $0<a\leq 25$. Analyzing these sets has led to several conjectures. One is that the set of logarithms of the elements of $Q_a(z)$ become normally distributed for any fixed a as $z\to\infty$. A second has to do with the prime divisors $p\leq z$ of the sets $Q_a(z)$. Clearly any prime divisor p of an element of $\mathcal{Q}_a(z)$ must have the property that $-a$ is a square modulo p. For such a p we might naively expect that approximately $2/p$ of the elements of $\mathcal{Q}_a(z)$ are divisible by p. Instead we conjecture that around $c_{p,a,z}/\sqrt{p}$ of the elements are divisible by p where $c_{p,a,z}$ is usually between 1 and 2.

KEYWORDS
Quadratic sieve; smooth numbers; Pell's equation

1. Introduction

In this article, we consider the sets of integers

$$\mathcal{Q}_a(z) := \{q : p|q^2+a \Rightarrow p\leq z\}.$$

If $a\neq 0$ then by Thue's theorem $\mathcal{Q}_a(z)$ is a finite set. When $a=\pm 1, \pm 2, \pm 4$ then for any z the entire set $\mathcal{Q}_a(z)$ can in principle be found by the Pell's equation method introduced by Størmer [Størmer 98]. For example, the complete set of 100-smooth numbers of the form q^2+1 has size 156. This is due to Luca [Luca 04]. Najman [Najman 10] found all of the 200-smooth numbers of the form q^2+1 and also considered the polynomials $q^2\pm 2$ and q^2+4 and used the Pell equation method to find all 200-smooth solutions and also all 100-smooth numbers of the form q^2-4. He remarks that the Pell equation method will not work for the polynomial q^2+a for other values of a than those already mentioned. See [Lehmer 64, Luca 04, Luca and Najman 11 and Najman 10] for details. For other values of a it is not clear how to generate the entire set, though by effective versions of Thue's theorem it can be done by a finite calculation.

We describe an algorithm which for each a and z leads to a subset $Q_a(z)\subset\mathcal{Q}_a(z)$ that can be calculated quickly and for which we conjecture that (for each a),

$$|Q_a(z)| = |\mathcal{Q}_a(z)|(1+o_a(1))$$

as $z\to\infty$. We use this algorithm to find $Q_a(500)$ for $1\leq a\leq 25$ and then we give some interesting statistical observations about these diophantine sets. The sets $\mathcal{Q}_a(z)$ are of interest in the study of the quadratic sieve factoring algorithm invented by Carl Pomerance [Pomerance 13].

In an earlier article [Conrey et al. 13] the authors together with Tara McLaughlin found a method to quickly generate many "z-smooth neighbors," numbers q and $q+1$ all of whose prime factors are smaller than some fixed number z. It was indicated in that paper how the method could be adapted to find many smooth values of $q(q+d)$ for any d. When $d=2a$ this is of course essentially the same problem as q^2-a^2.

For larger z and for all a our new quick method seems to find almost all z-smooth numbers of the form q^2+a. We have used our method for $0<|a|\leq 25$ and $z=500$, as mentioned earlier, and also for $a=1$ and $z=1900$ and for $a=2$ and $z=1000$. All of our data sets are available at aimath.org/∼conrey/smooth/morefiles.

Our algorithm is based on the identity:

$$(m^2+ax^2)(n^2+ay^2) = (mn-axy)^2+a(my+nx)^2$$

from which follows (with $x=y=1$)

CONTACT J. B. Conrey conrey@aimath.org American Institute of Mathematics, San Jose, CA 95112, USA.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

Proposition 1. *If m, n, and $\frac{mn-a}{m+n}$ are all integers then for any prime p, if $p|((\frac{mn-a}{m+n})^2+a)$ then $p|(m^2+a)$ or $p|(n^2+a)$.*

Proof. The proof is simply that

$$\left(\frac{mn-a}{m+n}\right)^2+a=\frac{(m^2+a)(n^2+a)}{(m+n)^2}.$$

□

This leads us to

Algorithm 1. Given a and z we first calculate

$$Q_a^{(0)}(z)=\{n\le z : p|n^2+a\Rightarrow p\le z\}.$$

Then we form a possibly larger set $Q_a^{(1)}(z)$ which includes $Q_a^{(0)}(z)$ together with any integer values of $\frac{q_1q_2-a}{q_1+q_2}$ with $q_1,q_2\in Q_a^{(0)}(z)$. We can repeat this process with $Q_a^{(1)}(z)$ in place of $Q_a^{(0)}(z)$ to form a new set $Q_a^{(2)}(z)$. We keep repeating this process until at some stage there are no new integers values found, i.e. we find an n such that $Q_a^{(n)}(z)=Q_a^{(n+1)}(z)$. Then we stop. The resulting set is $Q_a(z):=Q_a^{(n)}(z)$.

We have calculated $Q_a(500)$ for $0<|a|\le 25$ with $a\ne -b^2$; $Q_{-b^2}(100)$ with $1\le b\le 5$; and also $Q_1(1900)$ and $Q_2(1500)$.

It appears that for each a the set $\log Q_a(z)$ consisting of the logarithms of the elements of $Q_a(z)$ is approximately normal. We formalize this as

Conjecture 1. *There are functions $\mu_a(z)$ and $\sigma_a(z)$ such that*

$$\lim_{z\to\infty}\frac{\log Q_a(z)-\mu_a(z)}{\sigma_a(z)}=\mathcal{N}_{0,1},$$

where $\mathcal{N}_{0,1}$ denotes the unit normal distribution.

We don't have a guess for μ_a and σ_a but there is data about these in the tables in Section 3.

2. The operation behind the algorithm

It is convenient to describe our algorithm in terms of a group operation $\star_a$ which is defined by

$$m\star_a n=\frac{mn-a}{m+n}.$$

It can be shown that

Theorem 1. *We have*

$$\mathbb{N}\cap\langle Q_a^{(0)}(z)\rangle=\mathcal{Q}_a(z);$$

i.e. the positive integer points of the subgroup generated by our initial set $Q_a^{(0)}(z)$ with the group operation $\star_a$ are precisely the positive integers q for which $p|q^2+a\Rightarrow p\le z$.

In general it is too time and space consuming to compute very much of $\langle Q_a^{(0)}(z)\rangle$; so instead we compute a small part of it which seems to give most of the integer points.

Here is some data about the size of $Q_1(z)$ and its maximal element:

z	$\#Q_1(z)$	max
100	132	617427
200	621	1282794079
300	1666	1259851011582
400	3464	1259851011582
500	6544	36948955727316
600	10720	566334144961073
700	18369	1880980486194094
800	29657	122732491955797368
900	43292	258330078462753968
1000	58730	258330078462753968
1100	90726	328235377936173557
1200	119808	18590934165850666693
1300	176835	18590934165850666693
1400	216095	86412715207222970243
1500	281925	86412715207222970243
1600	315751	86412715207222970243
1700	433459	5558647451499052872645
1800	548835	5558647451499052872645
1900	646890	5558647451499052872645

The sets $Q_1(z)$ with $z=100,200,\ldots,1900$ may be found at aimath.org/∼conrey/smooth. From this set of data we make a conjecture about the size of $Q_1(z)$.

Conjecture 2. *There exists a $C>0$ such that for all $z\ge 1$ we have*

$$Q_1(z)\approx z\exp\left(C\sqrt{\log z}\right).$$

3. Other values of *a*

In this section we give some data about the sets $Q_a(500)$ for $1\le a\le 25$. Here max is the maximum element of $Q_a(500)$ and μ and σ denote the mean and standard deviation of the set of the logarithms of the elements of $Q_a(500)$.

a	$\#Q_a(500)$	max	μ	σ
1	6543	36948955727316	15.2306	4.51586
2	10123	740905937992184	16.4372	4.90819
3	11726	4174904929381219	16.9138	5.03408
4	11382	22542526183355414	16.8337	5.01502
5	11770	13494875248875220	16.8817	5.13242
6	17057	415466643146415876	17.7963	5.21811
7	14488	975303911197308	17.4934	5.14849
8	16504	57845217592272844	17.8410	5.36416
9	14072	1674200075341233	17.3869	5.18884
10	10520	16468480935656430	16.4151	4.89443

(*Continued*)

Continued.

a	$\#Q_a(500)$	max	μ	σ
11	20447	5488901165639322067	18.3829	5.52417
12	17055	8349809858762438	17.9331	5.34887
13	12634	658755374050997	16.8607	4.92614
14	21084	24175726919522264	18.3758	5.45911
15	15743	44577069068507912	17.8235	5.41516
16	15756	45085052366710828	17.7149	5.29102
17	18318	41539566273374107	17.9280	5.28586
18	14326	2955032783869080	17.5252	5.16796
19	18633	110140588909909729	18.0453	5.36788
20	18286	63096039891310453	18.0518	5.54245
21	15069	4182846397723553	17.5292	5.29204
22	11111	769288963903064	16.3031	4.56808
23	23647	63119562794014419	18.7992	5.62119
24	26846	830933286292831752	19.0590	5.64230
25	10576	2426342849673365	16.7899	4.93206

Here is some data about the size of $Q_a(500)$ and its maximal element for $a < 0$:

a	$\#Q_a(500)$	max	μ	σ
−2	3746	8626166844298	13.4452	4.05727
−3	5426	445886122971087	14.5636	4.34759
−5	6550	624466203267361	15.2899	4.56080
−6	6621	8858402990125534	15.2464	4.59277
−7	7674	4085476491878887	15.7339	4.73905
−8	7935	400382915634374	15.8532	4.73396
−10	8162	169112080417195	16.0585	4.92122
−11	8646	110559143357171	15.8697	4.68174
−12	10279	1046366255957646944	16.6000	4.96834
−13	8675	732422589726539	16.2163	4.98210
−14	9912	1964547833045108	16.2176	4.71113
−15	10224	362627003375442	16.4683	4.88602
−17	9146	189296371612887	16.1642	4.78417
−18	9904	8917538940735507	16.4156	4.89516
−19	8788	4531389322369616	16.1350	4.97972
−20	10363	1910765732910988	16.6810	4.94996
−21	10179	538628304240481	16.6300	5.07265
−22	14381	6810451517280656	17.4206	5.13343
−23	6704	24168393454607	15.1914	4.53338
−24	11388	17716805980251068	16.8831	5.10811

For the polynomials $q^2 - a$ when $a = b^2$ is a square, the sets of 500-smooth values are quite large; consequently, we only have data about 100-smooth values.

a	$\#Q_a(100)$	max	μ	σ
−1	16196	332110803172167361	16.0676	5.24944
−4	19047	664221606344334722	16.4259	5.30968
−9	21328	996332409516502083	16.6259	5.32210
−16	21892	1328443212688669444	16.7782	5.36145
−25	23243	1660554015860836805	16.8908	5.37247

Many of the solutions in this last table are imprimitive in the sense that they are small multiples of solutions of other problems. For example, odd smooth numbers of the form $q^2 - 1$ arise as twice smooth neighbors; i.e. if $m(m+1)$ is smooth then so is $2m(2m+2) = (2m+1)^2 - 1$. In our list above there are 2848 100-smooth even values of $q^2 - 1$; there are 2841 100-smooth odd values of $q^2 - 4$; there are 901 even 100-smooth values of $q^2 - 9$ with $3 \nmid q$; there are 2841 odd 100-smooth values of $q^2 - 16$; and 1290 even 100-smooth values of $q^2 - 25$ with $5 \nmid q$.

Conjecture 3. *There exists a $C_a > 0$ and a $z_a > 0$ such that for all $z \geq z_a$ we have*

$$Q_a(z) \approx z \exp\left(C_a \sqrt{\log z}\right).$$

4. Some graphics

Some histograms for the logarithms of the sets $Q_a(500)$ are given. These sets appear to be normally distributed which supports Conjecture 1.

5. What order of magnitude should we expect for the sets $Q_{a,z}$?

The likelihood that $m \star_a n$ is an integer is around $\frac{1}{m+n}$. If we sum over all m and n up to X we get

$$\sum_{m,n \leq X} \frac{1}{m+n} \sim x \log 4 > 1.38x.$$

So, for an initial set that is pretty dense we might expect that the first iteration might be around 38% larger.

6. The prime divisors of $q^2 + a$

Given a $q \in Q_a(z)$, how likely is it that $q^2 + a$ is divisible by a prime p? Clearly a necessary condition is that the Legendre symbol

$$\left(\frac{-a}{p}\right) = 1.$$

But given a p satisfying this condition, we might expect that such a p divides approximately $2/p$ of the numbers $q^2 + a$ for $q \in Q_a(z)$. The 2 is because there will be two solutions $\pm b$ to $b^2 + a \equiv 0 \bmod p$ and as q ranges over $Q_a(z)$ if it hits all residue classes modulo p equally often, both $q \equiv b \bmod p$ and $q \equiv -b \bmod p$ will lead to $q^2 + a \equiv 0 \bmod p$ and so the fraction of the time this holds will be $2/p$. In practice we find something quite different. It seems that p will divide $q^2 + a$ more than $1/\sqrt{p}$ of the time. The exact proportion seems difficult to pin down. Our data leads us to believe that it might be $c_{p,a,z}/\sqrt{p}$ for some constant $c_{p,a,z}$ between 1 and 2; but the data is also consistent with the proportion being as large as

$$\frac{\log \log p}{\sqrt{p}}.$$

The 6543 values of q for which $q^2 + 1$ is 500-smooth are distributed into residue classes modulo 11 as follows:

$p=11$	0	1	2	3	4	5	6	7	8	9	10
$a=1$	561	570	595	580	618	591	616	629	566	613	604

which is more or less uniformly distributed. But if we look at how the q for which q^2+1 is 500-smooth are distributed modulo 13 we get a different story:

$p=13$	0	1	2	3	4	5	6	7	8	9	10	11	12
$a=1$	316	334	313	321	335	1453	320	301	1556	335	306	338	315

The two solutions $q=5$ and $q=8$ of $q^2+1\equiv 0 \bmod 13$ have significantly larger, approximately equal, entries whereas the rest are evenly distributed but smaller. Of course, $q\equiv\pm 5 \bmod 13$ corresponds to $13|q^2+1$.

Here are the results modulo 13 for the distribution of $q^2+a \bmod p$ for some other values of a:

$p=13$	0	1	2	3	4	5	6	7	8	9	10	11	12
$a=2$	772	745	808	723	795	815	792	755	828	765	775	778	772
$a=3$	576	532	605	628	574	571	2669	2613	589	580	595	585	609
$a=4$	537	544	584	2626	557	572	575	542	595	596	2550	555	549
$a=5$	902	905	894	875	917	907	910	877	941	917	879	924	922
$a=6$	1309	1304	1307	1312	1355	1314	1362	1263	1270	1317	1335	1267	1342
$a=7$	1100	1129	1104	1142	1133	1096	1074	1118	1082	1154	1098	1134	1124
$a=8$	1258	1251	1227	1316	1310	1271	1246	1235	1265	1262	1313	1284	1266
$a=9$	707	728	3116	732	706	681	683	733	684	665	669	3277	691
$a=10$	506	513	533	508	2310	568	519	551	530	2357	532	513	580
$a=11$	1548	1563	1563	1540	1528	1596	1594	1579	1678	1531	1613	1581	1533
$a=12$	856	3783	788	880	872	840	908	846	839	834	845	874	3890
$a=13$	2886	798	798	854	824	797	827	820	810	764	817	807	832

We have analyzed data for all primes $p\le 500$ and the sets $Q_a(500)$. What we have found is that the q are uniformly distributed over all of the residue classes modulo p if $-a$ is not a square modulo p. We state this as

Conjecture 4. *If* $\left(\frac{-a}{p}\right)=-1$ *then the sets*

$$R_{a,p}(r,z):=\{q\in Q_a(z): q\equiv r \bmod p\}$$

are uniformly distributed in the sense that

$$\lim_{z\to\infty}\frac{|R_{a,p}(r_1,z)|}{|R_{a,p}(r_2,z)|}=1$$

for any choice of r_1 *and* r_2. *If* $\left(\frac{-a}{p}\right)=1$ *and either* $r_1^2\equiv r_2^2\equiv -a \bmod p$ *or* $r_1^2\not\equiv -a \bmod p$ *and* $r_2^2\not\equiv -a \bmod p$ *then this limit still holds.*

The second half of this conjecture states that if $-a$ is a square modulo p, say $b^2\equiv -a \bmod p$, then the q which are not $\pm b \bmod p$ will be equally likely to be in any other residue class and the q that are $\equiv b \bmod p$ are equally numerous (in the limit) as those that are $\equiv b \bmod p$. This likelihood of this latter event, that $q\equiv\pm b \bmod p$, seems to be larger by a factor which is about $\sqrt{p}$. For example, within the set $Q_1(1900)$ we find that 186252 of the 646890 elements q have q^2+1 divisible by 29. We define

$$c_{29,1,1900}=\sqrt{29}\times\frac{186252}{646890}=1.55\ldots$$

and in general

$$c_{p,a,z}:=\frac{|\{q\in Q_a(z): p|q^2+a\}|}{|Q_a(z)|}\sqrt{p}.$$

Here is some further data for $Q_1(1900)$:

p	5	13	17	29	37	41	53	61	73	89
$c_{p;1;1900}$	1.517	1.566	1.568	1.550	1.537	1.530	1.510	1.495	1.483	1.452
p	1733	1741	1753	1777	1789	1801	1861	1873	1877	1889
$c_{p;1;1900}$	0.969	0.973	0.978	0.979	0.968	1.00	0.970	0.955	0.987	0.969

Basically, the $c_{p,a,z}$ tend to decrease as p increases up to z.

Based on the idea that $c_{p,a,z}$ might be around 2 a lot of the time we have computed

$$r_a:=\prod_{\substack{5<p\le 500\\ \left(\frac{-a}{p}\right)=1}}\left(1-\frac{2}{\sqrt{p}}\right)^{-1}$$

and here we present a comparison of $|Q_a(500)|$ with r_a with $1\le a\le 25$;

a	1	2	3	4	5	6	7	8	9	10
r^a	5169	7129	20710	5169	7744	17474	5957	7129	5169	37668
Q^a	6543	10123	11726	11382	11770	17057	14488	16504	14072	10520
a	11	12	13	14	15	16	17	18	19	20
r^a	2355	20710	45142	4076	4158	5169	21816	7129	17911	7744
Q^a	20447	17055	12634	21084	15743	15756	18318	14326	18633	18286
a	21	22	23	24	25					
r^a	8014	21486	3363	17474	5169					
Q^a	15069	11111	23647	26846	10576					

It would be nice to have a precise conjecture about size of $Q_a(z)$ (Figure 1).

7. Summary

We have found a quick method to generate most of the finitely many z smooth solutions of q^2+a. This gives us a rich set of data with some interesting characteristics. This gives rise to many questions, foremost is to conjecture for each a an asymptotic formula for the number of z-smooth values of q^2+a as $z\to\infty$. Another question would be to guess a conjecture for each pair a, b

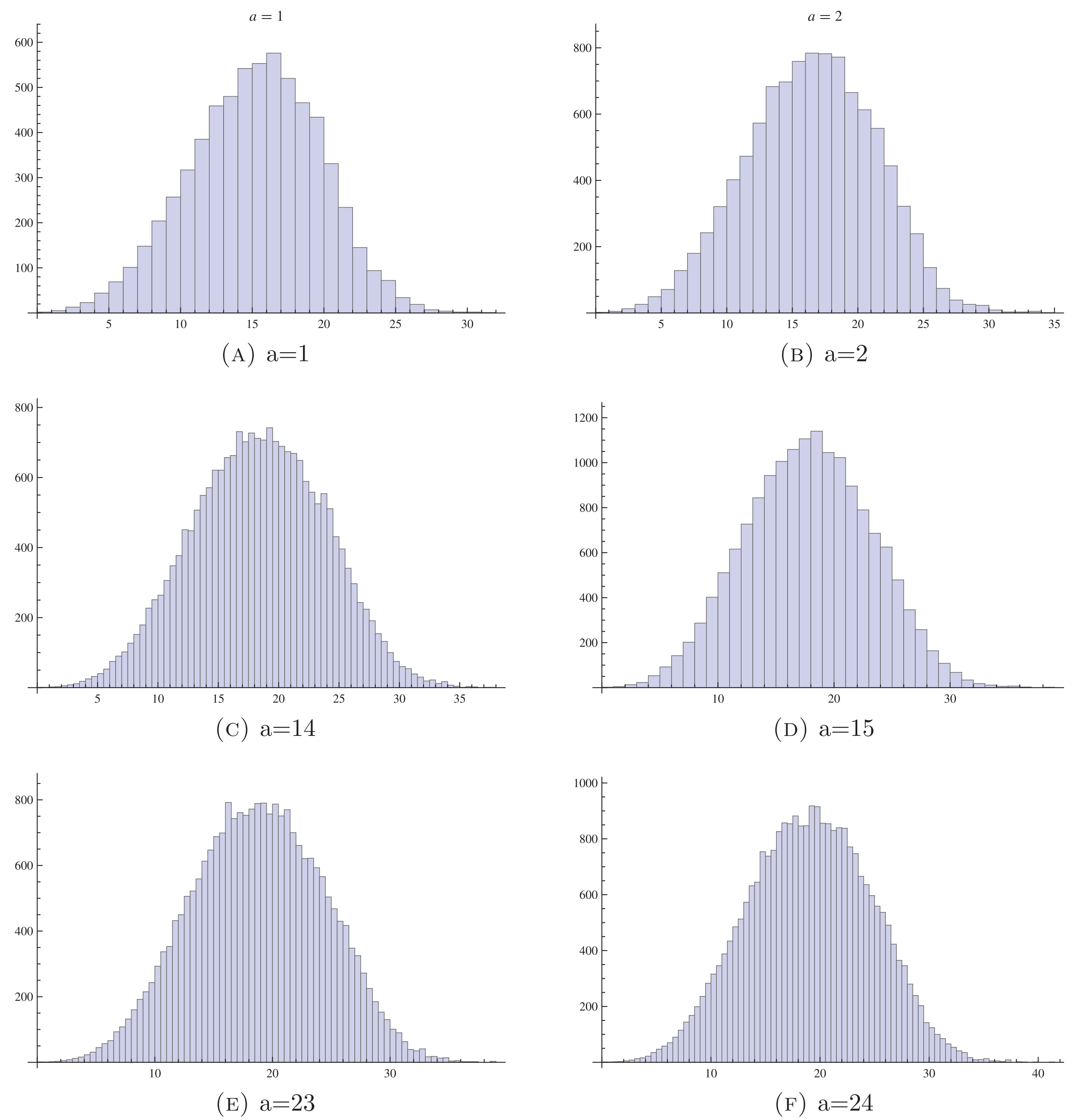

Figure 1. Histograms of $\log Q_a(500)$.

of the limit as $z \to \infty$ of the ratio of the number of z-smooth values of $q^2 + a$ to the number of z-smooth values of $q^2 + b$ (assuming this limit exists). For example the ratios

$$s_{2,1}(z) := \frac{\#Q_2(z)}{\#Q_1(z)}$$

are

z	100	200	300	400	500	600	700	800	900	1000
$s_{2;1}(z)$	1.735	1.320	1.28	1.251	1.547	1.883	2.031	1.611	1.654	1.784

It is not clear whether this is converging.

Funding

The first author was supported in part by a grant from the NSF.

References

[Conrey et al. 13] J. B. Conrey, M. A. Holmstrom, and T. L. McLaughlin. "Smooth Neighbors." *Exp. Math.* 22:2 (2013), 195–202.

[Lehmer 64] D. H. Lehmer. "On a Problem of Störmer." *Illinois J. Math.* 8 (1964), 57–79.

[Luca 04] F. Luca. "Primitive Divisors of Lucas Sequences and Prime Factors of x^2+1 and x^4+1." *Acta Acad. Paedagog. Agriensis Sect. Mat. (N.S.)* 31 (2004), 19–24.

[Luca and Najman 11] F. Luca and F. Najman. "On the Largest Prime Factor of x^2-1." *Math. Comp.* 80 (2011), 429–435.

[Najman 10] F. Najman. "Smooth Values of Some Quadratic Polynomials." *Glas. Mat. Ser. III* 45:2 (2010), 347–355.

[Pomerance 82] C. Pomerance. "Analysis and Comparison of Some Integer Factoring Algorithms." In *Computational Methods in Number Theory, Part I*, edited by H.W. Lenstra Jr. and R. Tijdeman, pp. 89–139. Amsterdam: Math Centre Tract, 1982.

[Størmer 98] C. Størmer. "Sur une Équation Indéterminée." *C. R. Paris* 127 (1898), 752–754.