In Praise of the

Bateman–Horn Conjecture

Alexander Zvonkin

(LaBRI — Université de Bordeaux)

Joint work with **Gareth Jones** (University of Southampton)
with computational assistance from **Jean Bétréma** (Bordeaux)

Seminar “Graphs on surfaces and curves over number fields”
Moscow, May 5, 2021

It [the Bateman–Horn conjecture] implies many known results, such as the prime number theorem and the Green–Tao theorem, along with many famous conjectures, such the twin prime conjecture and Landau's conjecture.

[. . .]

We hope to convince the reader that the Bateman–Horn conjecture deserves to be ranked among the Riemann hypothesis and abc-conjecture as one of the most important unproven conjectures in number theory.

From the paper "The Bateman–Horn conjecture: Heuristic, history, and applications", by S. L. Aletheia-Zomlefer, L. Fukshansky and S. R. Garcia, **2020**.

The conjecture belongs to the domain of Number theory. We have never heard of it before. Our way to this conjecture went through the classification of permutation groups **of prime degree**.

$*\qquad*\qquad*$

Publicity: Have a look at our new paper

Gareth A. Jones, Alexander K. Zvonkin
Klein's ten planar dessins of degree 11, and beyond (59 pages)

https://arxiv.org/pdf/2104.12015.pdf (less than two weeks ago)

Submitted to Фундаментальная и прикладная математика

Just in case, if there are young students in the audience:

degree — the number of points on which the group acts;
order — the number of elements in the group.

Symmetric group S_n:

degree $= n$,
order $= n!$.

Our interest is in the degree.

The groups of prime degree are few. . .

degree	2	3	4	5	6	7	8	9	10	11	12	13	14	15
#(groups)	1	2	5	5	16	7	50	34	45	8	301	9	63	104

degree	16	17	18	19	20	21	22	23	24	25
#(groups)	1954	10	983	8	1117	164	59	7	25 000	211

degree	26	27	28	29	30	31	32	33	34	35
#(groups)	96	2392	1854	8	5712	12	2 801 324	162	115	407

degree	36	37	38	39	40	41	42	43	44
#(groups)	121 279	11	76	306	315 842	10	9491	10	2113

degree	45	46	47	48
#(groups)	10 923	56	6	$\approx 195.8 \times 10^6$

The last result, for the degree 48, is dated 2020. Its author, Derek Holt, carried out the computation for several years. The result is not yet entirely verified. This is why Holt gives only an approximate value.

It is tempting to classify the groups of prime degree.

This work was started by Lagrange (1770), continued by Galois (1830), then by Burnside (1906) . . .

Finally, today, two and a half centuries after Lagrange and after the proof of the Mega-theorem of classification of the finite simple groups, the work may be considered as *almost* finished.

Why “almost”? — In fact, there still exists a fundamental question with an unknown answer. We will discuss it in a few minutes.

Remark. When I say that something is *unknown* I usually mean that there is, as yet, no proof. More often than not the true answer is “obvious” due to some indirect arguments and experimental results.

So. . .

The classification of permutation groups of prime degree goes as follows.

Case 1: Symmetric groups S_p and alternating groups A_p for p prime.

Clear, nothing to add.

Let $\mathsf{AGL}_1(p)$ be the one-dimensional affine group over $\mathbb{Z}_p$:

$$\mathsf{AGL}_1(p) = \{t \mapsto at + b \mid a, b \in \mathbb{Z}_p, a \neq 0\} = \mathsf{C}_p \rtimes \mathsf{C}_{p-1}.$$

Then

Case 2: The groups G such that $\mathsf{C}_p \leq G \leq \mathsf{AGL}_1(p)$:

$$G = \mathsf{C}_p \rtimes \mathsf{C}_d$$

where d is a divisor of $p - 1$, so that $\mathsf{C}_d \leq \mathsf{C}_{p-1}$.

Galois proved that these are the only solvable groups of prime degree.

Sporadic cases

The groups $\mathrm{PSL}_2(p) = \left\{ t \mapsto \dfrac{at+b}{ct+d} \mid a,b,c,d \in \mathbb{F}_p, ad-bc=1 \right\}$ act on $p+1$ points of the projective line $\mathbb{Z}_p \cup \{\infty\}$.

But there are three of them (also known to Galois) which can also act on p points:

Case 3a: The groups $\mathrm{PSL}_2(5)$, $\mathrm{PSL}_2(7)$ and $\mathrm{PSL}_2(11)$ acting on 5, 7 and 11 points, respectively.

Beside these three groups, there are two more groups:

Case 3b: Mathieu groups M_{11} and M_{23} acting of 11 and 23 points, respectively.

The most interesting (and difficult) case

Case 4: Let p be a prime, and $q = p^e$ be a prime power, $e \geq 1$. Let $\mathbb{F}_q$ be the finite field with q elements. Let $n \geq 2$. Then the groups G such that

$$\mathsf{PSL}_n(q) \leq G \leq \mathsf{P\Gamma L}_n(q)$$

act on

$$m = \frac{q^n - 1}{q - 1} = 1 + q + q^2 + \cdots + q^{n-1}$$

points of the projective space of dimension $n - 1$.

If it so happens that m is prime then the degree of G is prime.

We call such numbers $\boxed{m}$ projective primes.

Of course, for a given prime m it is easy to verify if it can be represented as

$$m = 1 + q + q^2 + \cdots + q^{n-1}$$

with $q = p^e$ a prime power.

Still, there is a largely open and quite fundamental question:

Open question: Are there infinitely many projective primes?
(Or, are there infinitely many projective groups of prime degree?)

A few examples

1. Fermat primes: $q = 2^{2^k}$ $(k = 0, 1, 2, 3, 4)$, $n = 2$

$$2^1 + 1 = 3, \quad 2^2 + 1 = 5, \quad 2^4 + 1 = 17, \quad 2^8 + 1 = 257, \quad 2^{16} + 1 = 65\,537.$$

Conjecture: There are no more Fermat primes.

2. Mersenne primes: $q = 2$, various n (51 examples are known)

$$2^2 - 1 = 3,\ 2^3 - 1 = 7,\ 2^5 - 1 = 31,\ 2^7 - 1 = 127,\ 2^{13} - 1 = 8191,\ \ldots$$

$$2^{82\,589\,933} - 1 \quad (24\,862\,048 \text{ digits}).$$

Conjecture: There are infinitely many Mersenne primes.

3. Just an example:

$$1 + 2^{59} + 2^{118} + \cdots + 2^{59\cdot 58} \;=\; \text{a prime with 1031 digits.}$$

Primality verification

In order to carry out experiments we need to undertake a primality verification on a large scale. What is the “practical complexity” of this task?

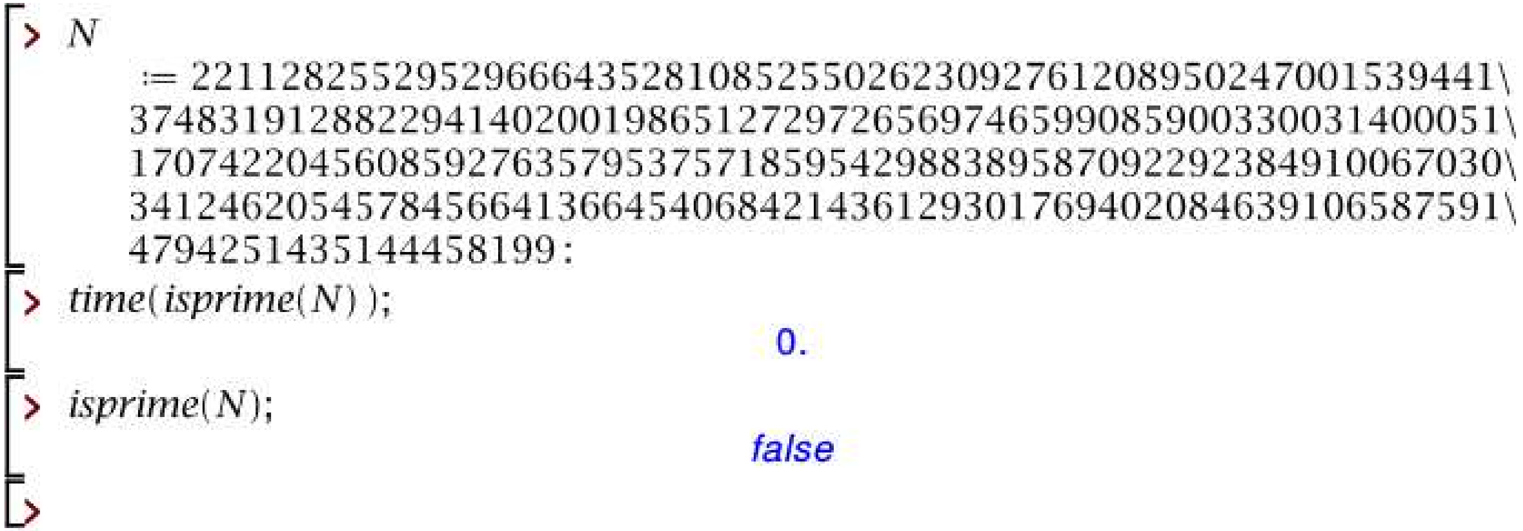
December 2009: a 232-digit number was successfully factored into a product of two 116-digit numbers. This result was the outcome of two years of work by a team of 13 researchers, and was crowned with a \$50 000 prize. The computation “time” is 4400 GHz-years.

February 2020, the current record: a 250-digit number is factored.

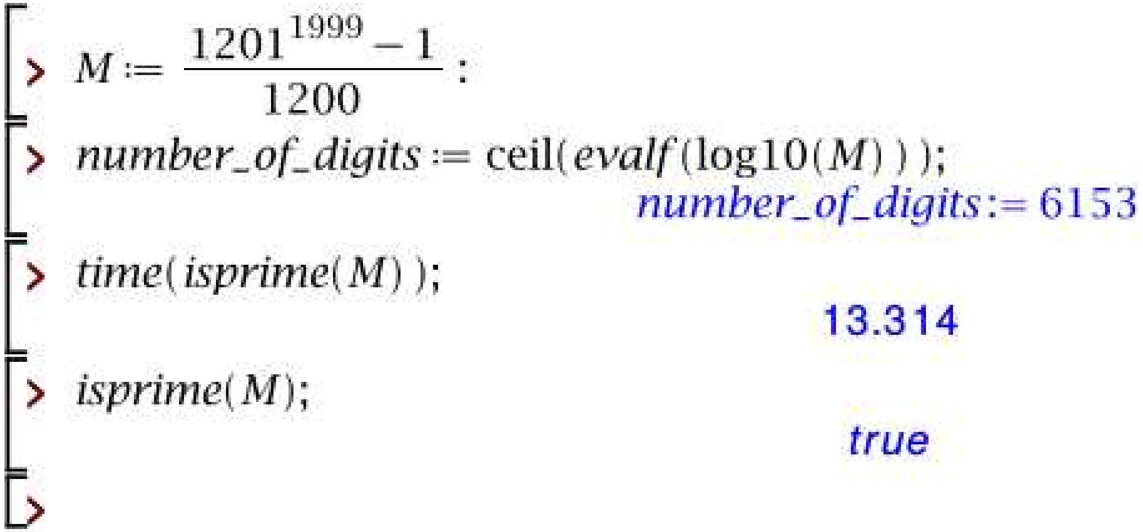
A 260-digit number is a current challenge: it still waits for its turn to be factored.

However, the verification of the fact that this 260-digit number is composite takes $<$ 0.0005 seconds on my laptop.
(I don’t know the exact time: Maple gives the CPU time within the accuracy 0.001 seconds, and it outputs 0.)

The hero is the **Rabin–Miller algorithm**.



```
> N
     := 221128255295296664352810852550262309276120895024700153944 1\
    374831912882294140200198651272972656974659908590033003140005 1\
    170742204560859276357953757185954298838958709229238491006703 0\
    341246205457845664136645406842143612930176940208463910658759 1\
    4794251435144458199:
> time(isprime(N));
                                 0.
> isprime(N);
                                false
>
```



$$> M := \frac{1201^{1999}-1}{1200}:$$

```
> number_of_digits := ceil(evalf(log10(M)));
                           number_of_digits := 6153
> time(isprime(M));
                                13.314
> isprime(M);
                                 true
>
```

One more example: **the Goormaghtigh conjecture** (1917): the Diophantine equation

$$\frac{x^n-1}{x-1}=\frac{y^k-1}{y-1}, \qquad n,k \geq 3, \qquad n \neq k$$

has only two solutions:

$$1+2+4+8+16=1+5+25=31$$

and

$$1+2+4+\cdots+2^{12}=1+90+90^2=8191.$$

Thus, 8191 is a projective prime for $(q,n)=(2,13)$; but 90 is not a prime power.

Conjecture: Only 31 is “doubly projective”. There are two different projective groups acting on 31 points: $\mathrm{PSL}_3(5)$ and $\mathrm{PSL}_5(2)$, and there are no other such examples. **Verified up to** 10^{18}.

I have mixed feelings concerning this conjecture: for about 20 years I thought that it was **my** conjecture.

Well... All the above was a starting point of our interest to prime values of polynomials. Here is a pioneering and really important but largely unknown conjecture.

Let $f(t) \in \mathbb{Z}[t]$ be a polynomial with integer coefficients. We would like it to have infinitely many prime values. There are three obvious necessary conditions:

1. The leading coefficient of f is positive.
2. f is irreducible over $\mathbb{Z}$.
3. The values of f do not have a common divisor > 1. (Another formulation: $f(t)$ is not identically zero modulo any prime.)

Examples that do not satisfy the 3rd condition:

- All the values of $f(t) = t^2 + t + 2 = t(t+1) + 2$ are even.

- All the values of $f(t) = t^9 - t^3 + 2520$ are divisible by 504.

Bunyakovsky conjecture (1857): The above three conditions are also sufficient. A polynomial satisfying conditions 1, 2, 3 takes prime values infinitely often.

The conjecture remains largely open. Even for $f(t) = t^2 + t + 1$ or $f(t) = t^2 + 1$ there is no proof in view.

Besides, there are 745 582 values of $t \le 10^7$ such that $t^2 + t + 1$ is prime, and 456 362 values such that $t^2 + 1$ is prime.

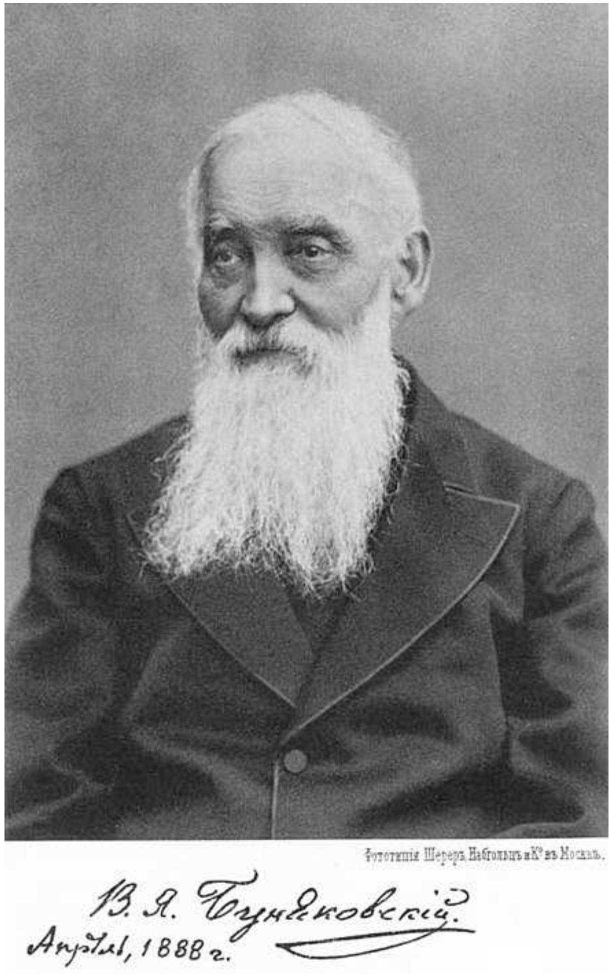
The only case which is proved is for polynomials of degree 1:

Theorem (Dirichlet, 1837): Let $a, b \in \mathbb{N}$ be coprime. Then the arithmetic progression $at + b$, $t \in \mathbb{N}$ contains infinitely many primes.

Example: There are infinitely many primes which terminate by $777\ldots7$ (17 times).

Proof: Take $a = 10^{17}$, $b = 777\ldots7$.

A short historical digression which will be interesting only to the Russian audience (it has nothing to do with our subject)



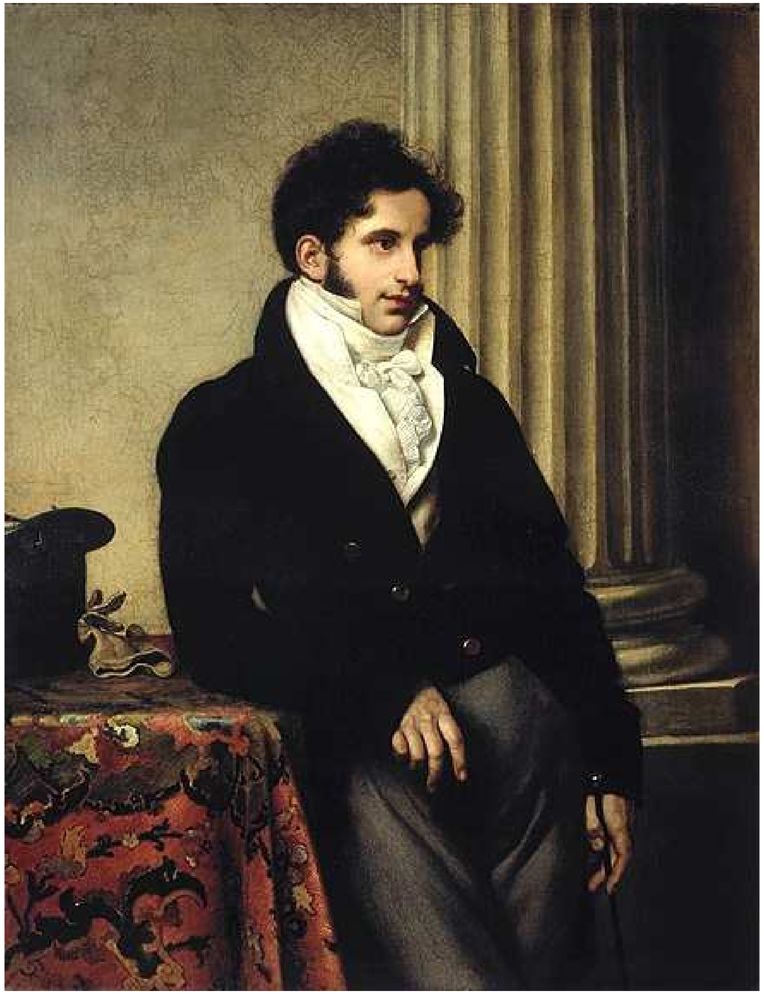
Viktor Yakovlevich Bunyakovsky (1804–1889)

Bunyakovsky was a student of Cauchy.

In Russia, he is known for the Cauchy–Bunyakovsky inequality which, in the Western tradition, is called after Cauchy–Schwarz.

Bunyakovsky proved it in 1859, Schwarz in 1888.

But there is more to say about Bunyakovsky.



Who is this young man (a portrait by Orest Kiprensky)?
I bet that nobody knows.

This is the **count Sergei Semionovich UVAROV** (1786–1855), a well-know personage of the Russian history of the XIX$^{\text{th}}$ century.

What do we know about him?

• He was the Minister of National Education (1833–49) under Nicholas I.

• He was a famous reactionary.

• As such, he was the author of the formula “Orthodoxy, Autocracy, Folk spirit” (Православие, самодержавие, народность).

Less known facts about him:

He was a renowned specialist in archaeology and in ancient Greek literature.

(His son was also an archaeologist, one of the founders of the Russian archaeological society and of the Moscow historical museum.)

Uvarov was on friendly terms with Goethe, with Alexander von Humboldt, Mme de Stael, Karamzin, and Zhukovsky.

He was elected an honorary member of the Russian Academy of Sciences in 1811 (at the age of 25 !), and was its president from 1818 until his death. He became minister in 1833.

Quoting: “While minister, Uvarov quietly promoted academic freedom and autonomy, raised academic standards, improved facilities, and opened higher education to the middle classes”. Finally he was fired by Nicholas.

(All this information is from the Wikipedia)

Beside that, Uvarov was also a great enthusiast of education in mathematics and physics.

He asked two academicians, Ostrogradsky and Bunyakovsky, to carry out a profound reform of mathematical education, namely:

- to write new textbooks for schools and universities;
- to enhance mathematical curricula of these institutions;
- the same for military and navy schools;
- to found pedagogical institutes in order to have enough teachers;
- to attract retired military officers to become science teachers;
- and so on.

Uvarov supported all their initiatives, and the reform was successful.

Thus, a high level of Russian mathematics at the end of XIX$^{\text{th}}$ and the beginningof XX$^{\text{th}}$ centuries is, in large part, due to these three persons: Uvarov, Ostrogradsky, and Bunyakovsky.

End of the digression

Concerning the groups of prime degree

For the purposes of group theory, we would need something less, and at the same time something more than the Bunyakovsky conjecture.

- It would be sufficient for us to have infinitely many prime values not for a single polynomial but for an infinite family of polynomials $1 + t + t^2 + \cdots + t^{n-1}$.

By the way, the exponent n must itself be prime, otherwise the above polynomial is reducible, like in the following example ($n = 6$):

$$1 + t + t^2 + t^3 + t^4 + t^5 = (1 + t)(1 + t^2 + t^4).$$

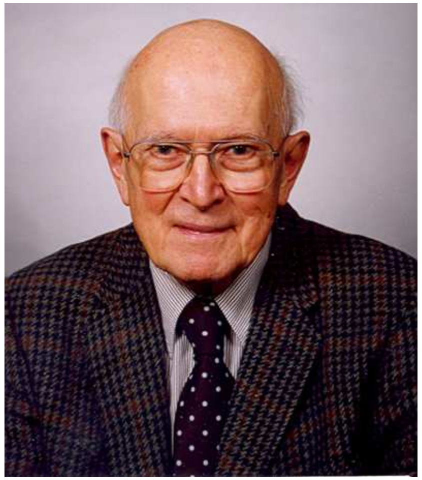
- On the other hand, we need prime values not at any t but at t being prime powers: $t = p^e$, $e \geq 1$.

There followed a series of generalizations and special cases of the Bunyakovsky conjecture:

- Generalized Bunyakovsky conjecture
- Dickson's conjecture (1904)
- Generalized Dickson's conjecture
- The Euler–Landau conjecture
- The Sophie Geramain conjecture
- Hardy and Littlewood's conjecture F (1923)
- Schinzel's hypothesis H (1960)
- $\cdots$

And finally —

The Bateman–Horn conjecture (1962)



Paul T. Bateman (1919–2012) and Roger A. Horn (born 1942)

Bateman at the time (1962) was a renowned specialist in number theory, and Horn was an undergraduate student who was able to write programs for the ILLIAC computer. Later he became a specialist in matrix analysis.

Notice that there was no Maple at their disposal and no other mathematical packages: everything had to be programmed from scratch.

Let $f_1, f_2, \ldots, f_k \in \mathbb{Z}[t]$ be polynomials with integer coefficients which satisfy the following conditions (similar to Bunyakovsky):

1. All f_i are indecomposable over $\mathbb{Z}$ (and hence coprime).

2. The leading coefficients of all of them are positive.

3. The product $f = f_1 f_2 \ldots f_k$ is not identically zero modulo any prime.

Let $Q(x)$ be the number of $t \leq x$ such that $f_i(t)$ **are ALL prime**.

Then $Q(x)$ is asymptotically equivalent to the following expression (see the next page). . .

I would like to begin not with the original Bateman–Horn conjecture but with its **improved version** proposed by Weixiong Li (2019).

Recall that $f(t)$ is the product $f = f_1 f_2 \ldots f_k$, and $Q(x)$ is the number of $t \leq x$ such that all $f_i(t)$ are prime. Then

$$\boxed{Q(x) \sim C(f) \int_a^x \frac{dt}{\prod_{i=1}^k \ln(f_i(t))}}$$

Here $C(f)$ is a constant factor to which I will return later, and the lower limit of integration, denoted here by a, should be adapted in such a way as to avoid the logarithmic singularities at $f_i(t) = 1$.

The original Bateman–Horn version:

Taking into account that if a polynomial $h = c \cdot t^d + \ldots$ then

$$\ln(h) \sim \ln(c) + d \cdot \ln(t) \sim d \cdot \ln(t)$$

we may write:

$$\boxed{Q(x) \sim \frac{C(f)}{d_1 \cdots d_k} \int_2^x \frac{dt}{\ln(t)^k}}$$

where $d_i = \deg f_i$.

I repeat the formula in order to discuss it:

$$Q(x) \sim \frac{C(f)}{d_1 \cdots d_k} \int_2^x \frac{dt}{\ln(t)^k}$$

When the leading coefficients of all the polynomials f_i are equal to 1 then both formulas give practically the same result. When some of the leading coefficients are greater than 1 the previous formula (that of Li) gives more accurate results.

Advantages of the present form:

1. No more trouble with the lower limit of integration: we just take $a = 2$.
2. The integral may be computed once and for all; certainly it was a very important advantage in 1962.
3. It becomes clear that, up to a multiplicative constant $C(f)$, the asymptotic behavior of $Q(x)$ depends only on the number of polynomials and of their degrees.

A further simplified version:

The equality

$$\int \frac{dt}{\ln(t)^k} = \frac{t}{\ln(t)^k} + k \cdot \int \frac{dt}{\ln(t)^{k+1}}$$

can be verified by differentiating both parts.

It implies

$$\int_2^x \frac{dt}{\ln(t)^k} \sim \frac{x}{\ln(x)^k}.$$

Advantage: can be computed on a pocket calculator, and sometimes even mentally (do you know that ln(20) is very close to 3?).

The difficult term is the constant $C(f)$. To follow. . .

$$C(f) = \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{\omega_f(p)}{p}\right)$$

where the product is taken over all primes p, and $\omega_f(p)$ is the number of solutions of $f(t) = 0$ mod p.

Thus, $C(f)$ is an infinite product. Its computation may present serious difficulties. In particular, it may be difficult to find $\omega_f(p)$.

A bit of interpretation:

1. $\left(1 - \dfrac{1}{p}\right)^k$ is the "probability" that a "randomly chosen" k-tuple of integers contains no integer divisible by p.

2. $1 - \dfrac{\omega_f(p)}{p}$ is the probability that $f(t)$ is not divisible by p. Since the function f is the product of f_i, this is the probability that no one of $f_i(t)$ is divisible by p.

Lemma: The above product converges to a constant $C > 0$.

A detailed proof takes seven pages. In the original paper by Bateman and Horn there are only a few hints.

Since the integral $\displaystyle\int_2^x \frac{dt}{\ln(t)^k}$ diverges when $x \to \infty$ we have the following corollary:

Corollary (Schinzel's hypothesis H): The polynomials $f_1, \ldots, f_k$ take prime values *simultaneously* infinitely many times.

The remaining part of the talk will mainly consist of

Examples

Example 1. Let us consider the simplest possible example: $k = 1$, so we have only one polynomial, and this polynomial is $f(t) = t$.

Then

$$Q(x) = \#(t \leq x) \text{ such that } t \text{ is prime.}$$

The equation $t = 0$ always has a unique solution modulo any p. Therefore

$$C(f) = \prod_p \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{1}{p}\right) = 1.$$

Conclusion: $Q(x) \sim \int_2^x \frac{dt}{\ln(t)} \sim \frac{x}{\ln(x)}$.

We recognize the Prime Number Theorem by Hadamard and de la Vallée Poussin (1896).

But look. . .

Relative errors of two estimates

x	$\dfrac{x}{\ln(x)}$	$\displaystyle\int_2^x \frac{dt}{\ln(t)}$
10^3	-13.69%	5.36%
10^4	-11.64%	1.30%
10^5	-9.45%	0.39%
10^6	-7.79%	0.16%
10^7	-6.64%	0.05%
10^8	-5.78%	0.013%
10^9	-5.10%	0.0033%
10^{10}	-4.56%	0.00068%
10^{11}	-4.13%	0.00028%
10^{12}	-3.77%	0.00010%
...	......	......
10^{25}	-1.77%	$3.12 \cdot 10^{-11}\%$

The number of primes up to 10^{25} was computed by J. Buethe, J. Franke, A. Jost, and T. Kleinjung (2013):

$$\begin{aligned} \text{true number} &= 176\,846\,309\,399\,143\,769\,411\,680 \\ \text{estimate} &= 176\,846\,309\,399\,198\,930\,392\,618 \end{aligned}$$

In fact, the integral formula

$$\pi(x) \sim \mathrm{Li}(x) = \int_2^x \frac{dt}{\ln(t)},$$

though being a particular case of Bateman–Horn, was invented long before them. Many sources attributed it to Gauss.

> Carl Friedrich Gauss considered the same question at age 15 or 16 “in the year 1792 or 1793”, according to his own recollection in 1849 (Gauss, Werke, 1863, Bd. 2, p. 444–447). (From Wikipedia)

And also in the same source:

> In a handwritten note on a reprint of his **1838** paper “*Sur l’usage des séries infinies dans la théorie des nombres*”, which he mailed to Gauss, **Dirichlet** conjectured $\langle\ldots\rangle$ that an even better approximation to $\pi(x)$ [as compared to the function $x/\ln(x)$] is given by the offset logarithmic integral function $\mathrm{Li}(x)$...

Remark on the accuracy of numeric integration

This question must be raised, especially since the intervals over which we integrate are extremely large.

(On the other hand, the function $\dfrac{1}{\ln(t)^k}$ varies very slowly, which is good for us.)

Unfortunately (or maybe fortunately) I don't know what is the algorithm of numeric integration used by Maple.

Then, what to do?

My approach was an experimental one. I made changes of variables in the integrals and computed them in various forms. Maple was not aware of the fact that it computed the same value.

The results were reassuring: the integrals may differ, for example, in the 15th digit.

Conjecture (Euler–Landau): There are infinitely many primes of the form $t^2 + 1$.

Euler — in a letter to Goldbach, 1752
Landau — at the International Congress of Mathematicians, 1912

Once again we have a single polynomial $f(t) = t^2+1$. The number of solutions of the equation $t^2 + 1 = 0 \bmod p$ is

$$\omega_f(p) = \begin{cases} 1 & p = 2 \\ 2 & p = 1 \bmod 4 \\ 0 & p = 3 \bmod 4 \end{cases}$$

A numerical estimate gives

$$C(f) = \prod_{p \le 10^8} \left(1 - \frac{1}{p}\right)\left(1 - \frac{\omega_f(p)}{p}\right) = 1.37281.$$

The relative error of the estimate

$$\frac{1.37281}{2} \int_2^x \frac{dt}{\ln(t)}$$

for $x = 10^9$ is 0.0085%.

Primes in arithmetic progressions

Let $a, b \in \mathbb{N}$ be coprime, and $f(t) = at + b$. Let us compute first the constant $C(f)$.

Recall that the Euler "totient function" $\varphi(n)$ is the number of integers less than n and coprime with n. We know that

$$\varphi(n) = n \cdot \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

The number of solutions of $at + b = 0 \bmod p$ is

$$\omega_f(p) = \begin{cases} 0 & \text{if } p|a \\ 1 & \text{if } \neg(p|a) \end{cases}$$

Therefore,

$$C(f) = \prod_{p} \left(1 - \frac{1}{p}\right)^{-1} \times \prod_{\neg(p|a)} \left(1 - \frac{1}{p}\right) = \prod_{p|a} \left(1 - \frac{1}{p}\right)^{-1} = \frac{a}{\varphi(a)}.$$

This is a rare case when the constant $C(f)$ may be found explicitly and not numerically.

The leading coefficient of the polynomial $at+b$ is $a>1$; therefore, we will integrate not $1/\ln(t)$ but $1/\ln(at+b)$.

The number of prime values of $at+b \le x \iff t \le (x-b)/a$ is approximated by

$$\frac{a}{\varphi(a)} \int_2^{(x-b)/a} \frac{dt}{\ln(at+b)}.$$

Making the change of variables

$$s = at+b, \quad ds = a \cdot dt, \quad dt = ds/a,$$

we get

$$\frac{a}{\varphi(a)} \int_2^{x} \frac{ds}{a \cdot \ln(s)} = \frac{1}{\varphi(a)} \int_2^{x} \frac{ds}{\ln(s)}.$$

The integral in the right-hand side is the approximation of the number of primes up to x.

Thus, our result may be interpreted as follows: there are $\varphi(a)$ integers b coprime with a, and for each of them the progression $at+b$ contains the part $1/\varphi(a)$ of primes.

Example: Take $a = 15$, $b = 11$, and $x = 10^8$. The number of primes $at + b \leq 10^8$ is $\underline{720\,313}$.

At the same time $\varphi(15) = 15\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right) = 8$, so that the above estimate gives

$$\frac{1}{8} \cdot \int_2^{10^8} \frac{ds}{\ln(s)} = \underline{720\,276.04}$$

Relative error -0.0051%.

Remark. Dirichlet proved only that the progression $at + b$ contains infinitely many primes. The above result is a theorem which Wikipedia attributes to **de la Vallée Poussin**, but without an explicit reference. Unfortunately, the papers by de la Vallée Poussin on this subject have hundreds of pages and are not available online.

An example taken from a paper by Amara, Devillers and Praeger on block designs

Let us take $f(t) = 32t^2 + 20t + 1$. In order to compute the constant $C(f)$ in front of the integral, we need to know $\omega_f(p)$ which is the number of solutions of the equation $f(t) = 0 \bmod (p)$.

- $p = 2$: $\quad 32t^2 + 20t + 1 = 0 \;\Leftrightarrow\; 1 = 0$: $\quad$ no solutions.

- $p \neq 2$: multiply $f(t)$ by 8:

  $$256t^2 + 80t + 8 = 256t^2 + 80t + 25 - 17 = (16t + 5)^2 - 17,$$

  so finally we have

  $$(16t + 5)^2 = 17 \bmod (p).$$

- $p = 17$: unique solution: $16t + 5 = 0$, hence $t = 5$.

- $p \neq 2, 17$: the equality $(16t+5)^2 = 17 \bmod p$ means that 17 is a *quadratic residue* modulo p.

Reminder: the Legendre symbol:

$$\left(\frac{q}{p}\right) = \begin{cases} 1 & \text{if } q \text{ is a quadratic residue mod } p, \\ -1 & \text{otherwise.} \end{cases}$$

The Gauss law of quadratic reciprocity:

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

In our case $q = 17$, hence $(q-1)/2 = 8$, hence $8 \cdot (p-1)/2 = 4(p-1)$ is even.

<u>Conclusion</u>: 17 is a quadratic residue modulo p if and only if p is a quadratic residue modulo 17.

Quadratic residues modulo 17 are: $1, 2, 4, 8, 9, 13, 15, 16$.

Finally we have

$$\omega_f(p) = \begin{cases} 0 & p = 2, \\ 1 & p = 17, \\ 2 & p \equiv 1, 2, 4, 8, 9, 13, 15, 16 \text{ mod } (17), p \neq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Computing the product below over $p \leq 10^8$ we get

$$C(f) = \prod_p \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{\omega_f(p)}{p}\right) = 4.721240276.$$

Since the leading coefficient of f is greater than 1, we will compute the estimate in the form

$$E(x) = C(f) \int_2^x \frac{dt}{\ln(f(t))}.$$

The results are summarized in the table below:

segment	#(prime $f(t)$)	$E(x)$	relative error
$t \leq 10^3$	326	314.49	−3.53%
$t \leq 10^4$	2421	2404.86	−0.67%
$t \leq 10^5$	19 394	19 438.26	0.23%
$t \leq 10^6$	162 877	163 182.75	0.19%
$t \leq 10^7$	1 405 448	1 406 630.14	0.084%
$t \leq 10^8$	12 357 532	12 362 961.06	0.044%

Beautiful !

At the same time, the above considerations demonstrate that the computation of the constant factor $C(f)$ may be rather intricate matter.

For example: what to do with cubic polynomials?

Twin primes

Let is take $f_1 = t$ and $f_2 = t + 2$. We want them to be simultaneously prime.

The equation $t(t+2) = 0 \bmod p$ has one solution for $p = 2$ and two solutions for all other primes p. Thus we have

$$\begin{aligned} C(f) &= \prod_p \left(1 - \frac{1}{p}\right)^{-2} \left(1 - \frac{\omega_f(p)}{p}\right) \\ &= 2 \prod_{p \geq 3} \left(1 - \frac{1}{p}\right)^{-2} \left(1 - \frac{2}{p}\right) = 1.320323630. \end{aligned}$$

For example, the number of pairs of twin primes up to 10^{18} is <u>808 675 888 577 436</u> (see the sequence A007508 in the Online Encyclopedia of Integer Sequences), while

$$1.320323630 \cdot \int_2^{10^{18}} \frac{dt}{\ln(t)^2} = \underline{808\,675\,900\,456\,220}.$$

The relative error is 0.000000147%. Don't forget that the infinitude of the twin primes is not yet proved!

Sophie Germain primes

Conjecture: There are infinitely many primes p such that $2p+1$ is also prime.

We take $f_1 = t$ and $f_2 = 2t+1$, so that $f(t) = t(2t+1)$. Solving the equation $f(t) = 0 \bmod p$ we find out that

$$\omega_f(p) = \begin{cases} 1 & p = 2 \\ 2 & p \geq 3 \end{cases}$$

A funny observation: $\omega_f(p)$ is exactly the same as in the case of twin primes. Therefore, the constant will also be the same: $C = 1.320323630$.

There is, however, one difference: the leading coefficient of the polynomial $f_2(t) = 2t+1$ is not 1. Therefore we will replace one of the $\dfrac{1}{\ln(t)}$ under the integral with $\dfrac{1}{\ln(f_2(t))}$. And, indeed, the estimation becomes better.

The number of Sophie Germain primes (that is, primes p such that $2p+1$ is also prime) is known up to $p \leq 10^{14}$: it is equal to 132 822 315 652 (see the sequence A092816 in the Online Encyclopedia of Integer Sequences). The estimate is

$$C \cdot \int_2^{10^{14}} \frac{dt}{\ln(t)\ln(2t+1)} = \underline{132\,822\,400\,361}.$$

The relative error is 0.000064%.

The conjecture, however, remains open: we still "don't know" if there are infinitely many primes like that.

Cunningham chains: $(p_1, p_2, \ldots, p_k)$ such that $p_{i+1} = 2p_i + 1$, and all of them are prime.

Example: $(89, 179, 359, 719, 1439, 2879)$.
The longest known Cunningham chain contains 19 terms.

The Bateman–Horn conjecture implies that there are infinitely many Cunningham chains of any given length. Just the integral diverges, that's all.

Green–Tao theorem

Quoting:

> “One of the most spectacular results in twenty-first century number theory is the Green–Tao theorem” (2004)

It is doubly spectacular since it is not a conjecture but a theorem.

Theorem: For each positive integer k, the prime numbers contain infinitely many arithmetic progressions of length k.

Example: 5, 11, 17, 23, 29 (length 5).

OK, we know now what to do. Let us take k polynomials:

$$f_1(t) = t, \quad f_2(t) = t + a, \quad \dots, \quad f_k(t) = t + (k-1)a,$$

so that their product is

$$f(t) = t(t+a)(t+2a)\cdots(t+(k-1)a).$$

$$f(t) = t(t+a)(t+2a)\cdots(t+(k-1)a).$$

Attention, danger! If $p \leq k$ and p does not divide a then $f(t)$ is identically zero mod p. Indeed, it is of degree k and has k distinct roots mod p, hence it vanishes for all elements of $\mathbb{Z}_p$.

Then let us take $a = \prod_{i=1}^{k} p_i$. For example, for $k = 5$, instead of $a = 6$ as in the previous example, we take $a = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 2310$.

Then we have

$$\omega_f(p) = \begin{cases} 1 & p = p_i, \ i \leq k \\ k & p = p_i, \ i > k \end{cases}$$

Thus

$$C(f) = \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)^{-k+1} \prod_{i=k+1}^{\infty} \left(1 - \frac{1}{p_i}\right)^{-k} \left(1 - \frac{k}{p_i}\right) > 0.$$

The theorem is proved!

Back to groups and “projective primes”

Reminder: a projective prime is a prime m of the form

$$m = 1 + q + q^2 + \cdots + q^{n-1}$$

where q is a prime power: $q = p^e$, $e \geq 1$. The exponent n must itself be prime, otherwise the polynomial $1 + t + \cdots + t^{n-1}$ would be reducible.

Jean Bétréma, using the program Julia, computed all projective primes $m \leq 10^{18}$. There are 1 974 311 of them.

Among them, there are:

- 1 974 010 numbers of the form $1 + p + p^2$ with p prime, and
- 301 projective primes of other types.

Taking $f_1 = t$ and $f_2 = 1 + t + t^2$ and computing the integral over $t \leq 10^9$ we get the Bateman–Horn estimate for the first number: 1 973 868. The relative error is 0.0072%.

Why there are so few projective primes of other types?

Let us take, for example, $m = 1 + p + p^2 + p^3 + p^4$, p prime. In order to have $m \leq 10^{18}$ **we must take the integral over** $t \leq 10^{18/4}$ **instead of** $t \leq 10^{18/2}$ as in the previous case.

The number of primes of this form is 252, the estimate gives 246.72.

Another example: the number of primes $m \leq 10^{18}$ of the form $m = 1+p^3+p^6$ is 10; the integral is taken over $[2, 10^{18/6}] = [2, 10^3]$; the ‘‘asymptotic estimate'' of this number is 12.06.

There is a single projective prime $m \leq 10^{18}$ for p prime and $n = 31$: it is $1+2+2^2+\cdots+2^{30}$. We look for $p \leq 10^{18/30} = 3.98$; thus, the only other candidate is $1+3+3^2+\cdots+3^{30}$, but it is composite. We did not try to get an asymptotic estimate of the number 1.

The constant factor

In order to compute the constant factor we need the following lemma.

Lemma. Let $f = t(1 + t + \cdots + t^{n-1})$, and consider the equation

$$f(t) = 0 \text{ mod } p.$$

Then:

$$\omega_f(p) = \begin{cases} 2, & p = n \text{ (namely, } t = 0 \text{ and } t = 1\text{)}, \\ n, & p \equiv 1 \text{ mod } n \text{ (namely, } t = 0 \text{ and } n - 1 \text{ primitive} \\ & \text{roots of unity of degree } n \text{ modulo } p\text{)}, \\ 1, & \text{otherwise (there is always the root } t = 0\text{)}. \end{cases}$$

For example, take $n = 7$, so that

$$f = t(1 + t + t^2 + t^3 + t^4 + t^5 + t^6),$$

and let $p = 43 \equiv 1$ mod 7. Then we have seven roots of f in $\mathbb{Z}_{43}$:

$$0; \quad 4,\ 4^2 = 16,\ 4^3 = 21,\ 4^4 = 41,\ 4^5 = 35,\ 4^6 = 11 \quad (4^7 = 1).$$

Conclusion

At the beginning of the talk, I cited a review paper by Aletheia-Zomlefer, Fukshansky and Garcia on the Bateman–Horn conjecture. The first version of this paper was called

One conjecture to rule them all: Bateman–Horn

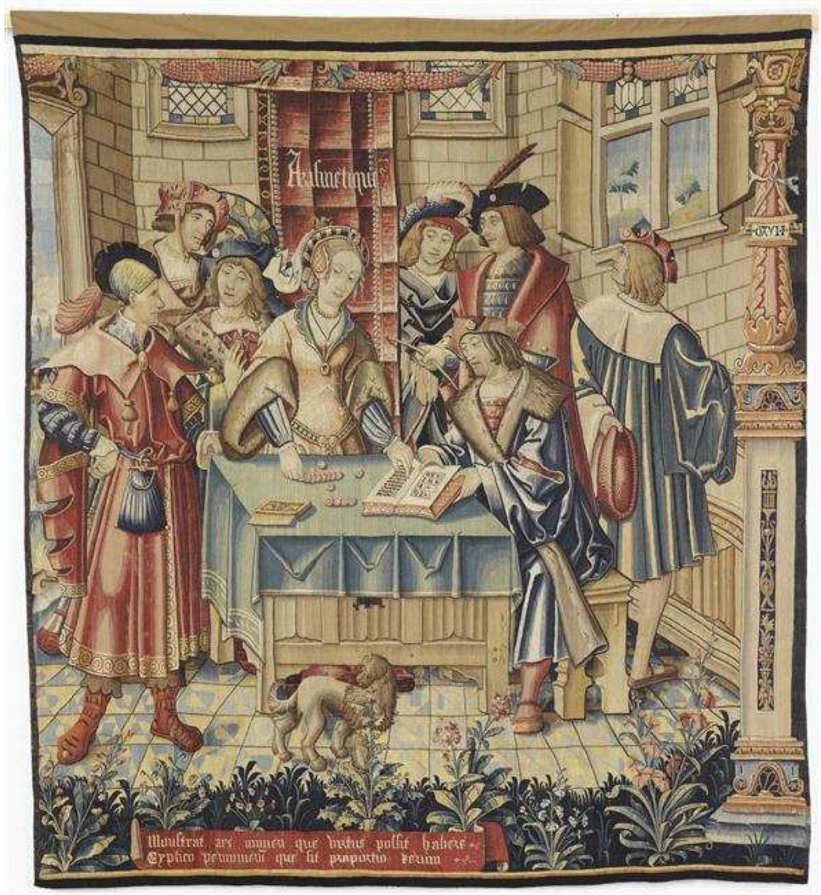
(An allusion to Tolkien: "One Ring to rule them all".)

And, indeed, we may continue indefinitely, inventing new and new conjectures *ad infinitum*.

Just two examples:

Our conjecture about projective primes (that there are infinitely many of them) does not figure in any known list of corollaries of the BH-conjecture. The same for polynomials coming from block designs.

Another remarkable feature of this conjecture is an **incredible accuracy** with which it predicts the number of "solutions", i. e., of prime values of polynomials, in all kinds of problems which fall into its framework.



L'arithmétique

Tapestry (around 1520) – Musée Cluny, Paris

The Latin inscription at the bottom:

Monstrat ars numeris que virtus possit habere

Explico pernumeru(m) que sit proportio rerum

The art of the number shows what virtue it may have:

I explain by the number which is the proportion of things

Thank you!