

A NOTE ON THE BATEMAN-HORN CONJECTURE

WEIXIONG LI

ABSTRACT. We report the results of our empirical investigations on the Bateman-Horn conjecture. This conjecture, in its commonly known form, produces rather large deviations when the polynomials involved are not monic. We propose a modified version of the conjecture which empirically demonstrates remarkable accuracy even for modest values of primes.

1. INTRODUCTION

In 1962, Bateman and Horn [1][2] proposed the following general conjecture concerning the distribution of primes generated from a set of polynomials¹.

Conjecture 1.1 (The Bateman-Horn Conjecture). Let $f_1, f_2, \dots, f_M \in \mathbb{Z}[x]$ be distinct irreducible polynomials with positive leading coefficients, and let

$$(1.1) \quad \pi(f_1, f_2, \dots, f_M; x) = \#\{n \leq x : f_1(n), f_2(n), \dots, f_M(n) \text{ are primes}\}$$

Suppose that $f(n) = \prod_{i=1}^M f_i(n)$ does not vanish identically modulo any prime, then for large values of x we have the following asymptotic expression

$$(1.2) \quad \pi(f_1, f_2, \dots, f_M; x) \sim \frac{C(f_1, f_2, \dots, f_M)}{\prod_{i=1}^M \deg f_i} \int_2^x \frac{dt}{(\log t)^M}$$

in which

$$(1.3) \quad C(f_1, f_2, \dots, f_M) = \prod_p \left(1 - \frac{1}{p}\right)^{-M} \left(1 - \frac{\omega_f(p)}{p}\right)$$

where the infinite product is over all primes p , and $\omega_f(p)$ is the number of solutions to the congruence equation $f(n) \equiv 0 \pmod{p}$.

The Bateman-Horn conjecture is very general, and many well-known conjectures, such as the Hardy-Littlewood Conjectures B, E, F, K, X, P [3][4], are all special cases of this conjecture.

In proposing this conjecture which now bears their names, Bateman and Horn provided the following heuristic arguments [1]. Since the probability of any number x being a prime number is roughly $\frac{1}{\log x}$, the probability of $f_i(n)$ being a prime number is $\frac{1}{\log f_i(n)}$, which for large n can be approximated by $\frac{1}{(\deg f_i) \log n}$, hence the expression in equation (1.2).

2010 *Mathematics Subject Classification*. Primary 11N32, 11N13, 11N05.

The author would like to thank William A. Stein and the Sage Development Team for making the software freely available to the public. This work would not have been possible without the Sage Mathematics Software.

¹The wording of the conjecture is mostly from [2].

Two questions naturally arise. (1) How good is the approximation of replacing $\frac{1}{\log f_i(n)}$ with $\frac{1}{(\deg f_i) \log n}$, and (2) how much better would the 'Bateman-Horn Conjecture' be without this approximation? To the best of this author's knowledge, there seems to be little discussion of these two questions in the literature [2].

Let us perform a naive estimation to answer the first question. Let $f_i(n) = a_i n^{m_i} + b_i n^{m_i-1} + \dots$, then $\log f_i(n) = \log(a_i n^{m_i}) + \log(1 + o(\frac{1}{n})) = m_i \log n + \log a_i + o(\frac{1}{n})$. we therefore see that when the polynomial is not monic, the relative errors caused by ignoring the term $\log a_i$ could be potentially significant for typical counting values around $n \sim 10^{15}$. It is interesting to note that historically most of empirical studies have been performed on monic polynomials where $\log a_i = 0$, therefore the errors never showed up in those studies [5][6]. The Hardy-Littlewood Conjecture F was expressed as a function of primes rather than the variable n [3][4], therefore the issue with non-monic polynomials is avoided.

In the next section, we will propose a modified version of the Bateman-Horn conjecture, we then perform empirical computations to compare the two versions of the conjecture with our numerical results.

2. EMPIRICAL STUDY

2.1. The Modified Conjecture. We will propose the following modified conjecture.

Conjecture 2.1 (The Modified Bateman-Horn Conjecture). Let $f_1, f_2, \dots, f_M \in \mathbb{Z}[x]$ be distinct irreducible polynomials with positive leading coefficients, and let

$$(2.1) \quad \pi(f_1, f_2, \dots, f_M; x) = \#\{n \leq x : f_1(n), f_2(n), \dots, f_M(n) \text{ are primes}\}$$

Suppose that $f(n) = \prod_{i=1}^M f_i(n)$ does not vanish identically modulo any prime, and n_0 is the smallest integer such that $\forall n > n_0, f_i(n) > 1, 1 \leq i \leq M$, then for large values of x we have the following asymptotic expression

$$(2.2) \quad \pi(f_1, f_2, \dots, f_M; x) \sim C(f_1, f_2, \dots, f_M) \int_{n_0}^x \frac{dt}{\prod_{i=1}^M \log f_i(t)}$$

where $C(f_1, f_2, \dots, f_M)$ is given by equation (1.3).

We now apply this modified conjecture to two cases involving non-monic polynomials.

2.2. Sophie Germain Primes. Application of the Bateman-Horn Conjecture to Sophie Germain primes was briefly discussed in [2], but the authors made no mention of its accuracy. A prime p is a Sophie Germain prime if $2p + 1$ is also a prime. Therefore we have two polynomials

$$\begin{aligned} f_1(n) &= n \\ f_2(n) &= 2n + 1 \end{aligned}$$

For any prime $p > 2$, the congruence equation $f(n) = n(2n + 1) \equiv 0 \pmod{p}$ has two solutions, corresponding to $n \equiv 0 \pmod{p}$ and $n \equiv \frac{p-1}{2} \pmod{p}$. For $p = 2$, $f(n) = n(2n + 1) \equiv 0 \pmod{p}$ has only one solution $n \equiv 0 \pmod{p}$. Therefore the Bateman-Horn constant C in equation (1.3) is given by

TABLE 1. $\pi_{SG}(x)$: Actual versus Conjectures

x	$\pi_{SG}(x)$	$2C_2 \int_2^x \frac{dt}{(\log t) \log(2t+1)}$	$2C_2 \int_2^x \frac{dt}{(\log t)^2}$
10^2	10	10	14
10^3	37	39	46
10^4	190	195	214
10^5	1171	1166	1249
10^6	7746	7811	8248
10^7	56032	56128	58754
10^8	423140	423294	440368
10^9	3308859	3307888	3425308
10^{10}	26569515	26568824	27411417

$$C = 2 \prod_{p>2} \frac{p(p-2)}{(p-1)^2} = 2C_2$$

where $C_2 = 0.66016181584686957393$ [7] is the Hardy-Littlewood twin-prime constant [3][4]. Therefore equation (2.2) becomes

$$(2.3) \quad \pi_{SG}(x) \sim 2C_2 \int_2^x \frac{dt}{(\log t) \log(2t+1)}$$

We generated about 29 million Sophie Germain primes using the freely available mathematical software SageMath v8.3 [8]. The largest Sophie Germain prime in the set is of the order 1.1×10^{10} . The program uses the prime set generated by the function Primes(), and for every prime p in the prime set, it tests whether $2p+1$ is also in the prime set. Table (1) lists our results. The column with the heading $\pi_{SG}(x)$ corresponds to the actual counting from our empirical computations, while the next two columns correspond to the modified Bateman-Horn conjecture and the originally proposed conjecture, respectively. As can be clearly seen, the modified conjecture gives remarkably close results compared with the actual counting, while the original version of the conjecture produces rather significant deviations. We point out that the modified version of the conjecture gives remarkably accurate estimates for even very modest values of x . That is, equation (2.2) is not just an asymptotic expression, it seems to give very close estimates for almost all values of x .

2.3. $f(n) = 6n^2+1$. We now present empirical data on the polynomial $f(n) = 6n^2+1$. Denote the prime counting function by $\pi_6(x)$. We need to calculate $\omega(p)$, the number of solutions to the congruence equation $6n^2+1 \equiv 0 \pmod{p}$. For primes $p = 2, 3$, we have $\omega(2) = \omega(3) = 0$. For any prime $p > 3$, note that this polynomial has discriminant $D = -24$, therefore $\omega(p) = 1 + (-24|p)$ where $(-24|p)$ is the Legendre symbol. The Bateman-Horn constant in equation (1.3) is then

$$(2.4) \quad C = 3 \prod_{p>3} \left(1 - \frac{(-24|p)}{p-1} \right)$$

which is essentially the expression given by the Hardy-Littlewood Conjecture F [3][4]. This infinite product is known to converge extremely slowly (and not absolutely). One

TABLE 2. $\pi_6(x)$: Actual versus Conjectures

x	$\pi_6(x)$	$C \int_1^x \frac{dt}{\log(6t^2+1)}$	$\frac{C}{2} \int_1^x \frac{dt}{\log t}$
10^2	27	25	31
10^3	155	162	189
10^4	1176	1195	1332
10^5	9445	9469	10299
10^6	78422	78514	84096
10^7	671361	670963	711171
10^8	5859476	5859288	6163042
10^9	52007341	52009622	54386431

technique² to accelerate its convergence is to multiply the following identity on both sides of equation (2.4):

$$(2.5) \quad L(1, (-24 | \cdot)) = \prod_{p>3} \left(1 - \frac{(-24 | p)}{p} \right)^{-1}$$

We then have

$$(2.6) \quad C = \frac{3}{L(1, (-24 | \cdot))} \prod_{p>3} \left(\frac{1 - \frac{(-24 | p)}{p-1}}{1 - \frac{(-24 | p)}{p}} \right)$$

The infinite product in equation (2.6) converges as p^{-2} . We note that the character $(-24 | \cdot)$ is primitive with a Conrey notation $\chi_{24}(5, \cdot)$ in the online L -function database [9], and the value of its corresponding Dirichlet L -function at $s = 1$ has a closed form, $L(1, (-24 | \cdot)) = \pi/\sqrt{6}$. Truncating the infinite product in equation (2.6) up to primes $\sim 10^6$, we get $C = 2.139124879$.

We generated 148 million primes of the form $6n^2 + 1$ using the mathematical software SageMath v8.3. The primality test is performed using the built-in function `is_prime()`. The largest prime thus generated has the order of magnitude of 5.4×10^{19} , and the corresponding value for $n \approx 7.3 \times 10^9$. Table (2) lists the value of $\pi_6(x)$ and compares it with the estimates from the modified version of the conjecture as well as the original version of the conjecture. Again one can clearly see the remarkable accuracy of the modified conjecture even for modest values of x , while the original version of the conjecture shows considerable deviations, an indication that the value of $x \sim 10^9$ is too small for the asymptotic approximation to be accurate.

3. CONCLUSION

Empirical computations suggest that the modified Bateman-Horn conjecture, equation (2.2), to be quite accurate over a much wider range of prime values, which is strikingly similar to the fact that $\text{Li}(x)$ is a much better approximation of $\pi(x)$ than the asymptotic expression $\frac{x}{\log x}$ in the classic Prime Number Theorem. Further work

²This technique was suggested by K. Conrad on a Mathoverflow discussion forum: <https://mathoverflow.net/questions/31150/calculating-the-infinite-product-from-the-hardy-littlewood-conjecture-f>

is still needed, such as an estimate of error bounds, as well as possible adaptation to other problems such as the Mersenne primes.

REFERENCES

1. P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. 16 (1962), 363-367. MR 0148632
2. S. L. Aletheia-Zomlefer, L. Fukshansky, S. R. Garcia, *The Bateman-Horn Conjecture: Heuristics, History, and Applications*, 2019. <https://arxiv.org/abs/1807.08899>
3. 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 Mathematica, 44:1-70, 1922.
4. Projecteuclid: <https://projecteuclid.org/euclid.acta/1485887559>
5. D. Shanks. *On the conjecture of Hardy and Littlewood concerning the number of primes of the form $n^2 + a$* . Math. Comp, 14:321-332, 1960.
6. D. Shanks, M. Lal, *Bateman's Constants Reconsidered and the Distribution of Cubic Residues*, *Mathematics of Computation*, Vol. 28, No. 117, January 1972.
7. G. Niklasch, <https://oeis.org/A001692/a001692.html>, May 2019.
8. William A. Stein et al. Sage Mathematics Software (Version 8.3), The Sage Development Team, 2018, <http://www.sagemath.org>.
9. LMFDB, <http://www.lmfdb.org/>, May, 2019.

3813 PATTY BERG CT, WOODRIDGE, IL 60517

E-mail address: `weixiongli@comcast.net`