Binomial Probability of Prime Number of Successes

Sungjin Kim, Nilotpal Kanti Sinha

California State University Northridge
Department of Mathematics
707107@gmail.com

Dec 17, 2019

Introduction

- Sums of binomial coefficients for $n \in \mathbb{N}$,

$$\sum_{k\geq 0}\binom{n}{k} = 2^n, \quad \sum_{k\geq 0}\binom{n}{2k} = 2^{n-1}.$$

Introduction

- Sums of binomial coefficients for $n \in \mathbb{N}$,

$$\sum_{k\geq 0}\binom{n}{k} = 2^n, \quad \sum_{k\geq 0}\binom{n}{2k} = 2^{n-1}.$$

- Generally,

$$\sum_{k\geq 0}\binom{n}{ak+b} = \frac{1}{a}\sum_{k=1}^{a}\omega^{-bk}(1+\omega^k)^n = \frac{2^n}{a}\sum_{k=1}^{a}\cos^n\frac{k\pi}{a}\cos\frac{(n-2b)k\pi}{a}$$

where $0 \leq b < a, n \geq 0$ and $\omega = e^{2\pi i/a}$ is a primitive a-th root of unity.

Question

• (A Mathematics Stack Exchange (MSE) question by N. K. Sinha) What is known about asymptotic order/lower and upper bound of the sum

$$S_n = \sum_{p \le n} \binom{n}{p},$$

where the sum is over all prime numbers $p \le n$.

Question

• (A Mathematics Stack Exchange (MSE) question by N. K. Sinha) What is known about asymptotic order/lower and upper bound of the sum

$$S_n = \sum_{p \le n} \binom{n}{p},$$

where the sum is over all prime numbers $p \le n$.
Then $S_n/2^n$ is the probability that the number of successes in n independent Bernoulli trials is a prime number, where the probability of success in each trial is $1/2$.

Notations

• $\mathbf{P}(A)$ is the probability of the event A.
• $T_n \sim \mathrm{B}(n, \frac{1}{2})$ is the binomial distribution with n trials, and the probability of success is $1/2$. Then we have $\mathbf{P}(T_n = k) = \binom{n}{k}/2^n$ for $0 \le k \le n$. T_n has the mean $n/2$, and the standard deviation $\sqrt{n}/2$.
• $\mathcal{P}$ is the set of prime numbers. Thus, $\mathbf{P}(T_n \in \mathcal{P}) = S_n/2^n$.
• $\pi(y) = \sum_{p \le y} 1$ is the number of primes not exceeding y.
• $A(n) \ll B(n)$ means $|A(n)| \le cB(n)$ for some positive absolute constant c.
• $\binom{x}{v} = \frac{\Gamma(x+1)}{\Gamma(v+1)\Gamma(x-v+1)} = \binom{x}{x-v}$ is the extension of binomial coefficients for real $x > 0$ and $v \ge 0$. For any $0 \le v_1 \le v_2 \le x/2 \le v_3 \le v_4 \le x$, we have $1 \le \binom{x}{v_1} \le \binom{x}{v_2} \le \binom{x}{x/2} \ge \binom{x}{v_3} \ge \binom{x}{v_4} \ge 1$.
• $S_x = \sum_{p \le x} \binom{x}{p}$ is an extension of S_n to positive real numbers.
• The letters j, k, n, p are integers. In particular, p denotes a prime. The letters $\alpha, \beta, \epsilon, t, v, x, X$ are real numbers. We write $c_1, c_2, \dots$ for absolute positive constants that can be effectively computed.

Initial comments and an answer on MSE

• (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.

Initial comments and an answer on MSE

• (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.

• (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.

Initial comments and an answer on MSE

• (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.

• (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.

• (N. K. Sinha) @QiaochuYuan. I think this approach will work. The square root in the denominator looks a little doubtful.

Initial comments and an answer on MSE

• (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.

• (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.

• (N. K. Sinha) @QiaochuYuan. I think this approach will work. The square root in the denominator looks a little doubtful.

• (Michael Lugo) Experimentally $O(2^n/(logn))$ looks more likely than the form with the square root.

Initial comments and an answer on MSE

- (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.
- (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.
- (N. K. Sinha) @QiaochuYuan. I think this approach will work. The square root in the denominator looks a little doubtful.
- (Michael Lugo) Experimentally $O(2^n/(logn))$ looks more likely than the form with the square root.
- (i707107) Answered with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n \log\log n}{\log n}$ where the lower bound is a conjecture.

Initial comments and an answer on MSE

• (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.

• (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.

• (N. K. Sinha) @QiaochuYuan. I think this approach will work. The square root in the denominator looks a little doubtful.

• (Michael Lugo) Experimentally $O(2^n/(logn))$ looks more likely than the form with the square root.

• (i707107) Answered with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n \log\log n}{\log n}$ where the lower bound is a conjecture.

• (i707107) Improved the answer with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n\sqrt{\log\log n}}{\log n}$ where the lower bound is a conjecture.

Initial comments and an answer on MSE

- (Qiaochu Yuan) The asymptotics should be dominated by the distribution of primes near $n/2$. By the Prime Number Theorem (PNT) we expect such primes to locally have a density of about $\frac{1}{\log(n/2)}$. So a rough guess for the asymptotics is $O(\frac{1}{logn}\binom{n}{n/2})$, which works out to something like $O(2^n/(\sqrt{n}\log n))$. This shouldn't be hard to get experimental data on.
- (i707107) @QiaochuYuan According to your guess, $O(2^n/logn)$ is plausible since you began with the density.
- (N. K. Sinha) @QiaochuYuan. I think this approach will work. The square root in the denominator looks a little doubtful.
- (Michael Lugo) Experimentally $O(2^n/(logn))$ looks more likely than the form with the square root.
- (i707107) Answered with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n \log\log n}{\log n}$ where the lower bound is a conjecture.
- (i707107) Improved the answer with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n\sqrt{\log\log n}}{\log n}$ where the lower bound is a conjecture.
- (i707107) Improved the answer with $\frac{2^n}{\log n} \ll S_n \ll \frac{2^n}{\log n}$ where the lower bound is a conjecture.

Numerical observations on $(n, (S_n \log(n/2))/2^n)$ for large n's-by N. K. Sinha

n	$S_n \log(n/2)/2^n$	n	$S_n \log(n/2)/2^n$
100000	1.069169869	250000	0.986114371
110000	0.94301485	260000	0.965609639
120000	0.917190017	270000	0.973894862
130000	1.009817376	280000	0.99483856
140000	1.027465936	290000	0.953542586
150000	0.974742038	300000	1.028188428
160000	1.029105385	310000	0.993445284
170000	0.965422147	320000	1.017001058
180000	1.119848774	330000	0.869868372
190000	1.054380578	340000	1.073959735
200000	0.948608301	350000	0.873428088
210000	0.972819167	360000	1.090734815
220000	0.904355813	370000	1.024869577
230000	0.973834543	380000	0.965571714
240000	1.039784878	390000	1.025289725

Main results

Theorem (1)

There is an absolute constant $c_0 > 0$ such that for almost all n,

$$S_n = \frac{2^n}{\log(n/2)} + O(2^n e^{-c_0 (\log n)^{1/3}/(\log\log n)^{1/3}}) \text{ as } n \to \infty.$$

Here, almost all means that the number of $n \in [1, N] \cap \mathbb{Z}$ for which the asymptotic formula fails is $O(Ne^{-c_0(\log N)^{1/3}/(\log\log N)^{1/3}})$.

Theorem (2)

We have

$$\alpha := \liminf_{n\to\infty} \frac{S_n \log n}{2^n} \le 1 \le \limsup_{n\to\infty} \frac{S_n \log n}{2^n} \le 4.$$

Main results

Theorem (3)

The statement $\alpha > 0$ holds if and only if there are constants $b_1, b_2 > 0$ such that

$$\pi\left(\frac{n}{2} + b_1\sqrt{n}\right) - \pi\left(\frac{n}{2} - b_1\sqrt{n}\right) \geq \frac{b_2\sqrt{n}}{\log n} \text{ for all } n \geq N_0(b_1, b_2).$$

Proof of Theorem 1

Lemma (Hoeffding's Inequality)

Let $X_1, \ldots, X_n$ be independent bounded random variables with $a \le X_i \le b$ for all i, and $\overline{X} = \frac{1}{n}\sum X_i$. Then for all $t \ge 0$,

$$\mathbf{P}\left(\left|\overline{X} - \mathbf{E}(\overline{X})\right| \ge t\right) \le 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

As a consequence, we have for sufficiently large real $x > 0$, $r > 0$, and $B_x = \left\{k \le x : \left|k - \frac{x}{2}\right| \ge r\sqrt{x}\right\}$,

$$\frac{1}{2^x}\sum_{k \in \mathcal{P} \cap B_x} \binom{x}{k} \le \frac{1}{2^x}\sum_{k \in B_x} \binom{x}{k} \le 4e^{-2r^2}.$$

Proof of Theorem 1

Lemma (Hoeffding's Inequality)

Let $X_1, \ldots, X_n$ be independent bounded random variables with $a \le X_i \le b$ for all i, and $\overline{X} = \frac{1}{n}\sum X_i$. Then for all $t \ge 0$,

$$\mathbf{P}\left(\left|\overline{X} - \mathbf{E}(\overline{X})\right| \ge t\right) \le 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

As a consequence, we have for sufficiently large real $x > 0$, $r > 0$, and $B_x = \left\{k \le x : \left|k - \frac{x}{2}\right| \ge r\sqrt{x}\right\}$,

$$\frac{1}{2^x}\sum_{k \in \mathcal{P} \cap B_x} \binom{x}{k} \le \frac{1}{2^x}\sum_{k \in B_x} \binom{x}{k} \le 4e^{-2r^2}.$$

Here, r may depend on x.

Main ingredients

We need Stirling's formula of the form:

$$\Gamma(x+1) = \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \left(1 + O\left(\frac{1}{x}\right)\right).$$

The following is the zero density estimate by Huxley [H].

Lemma (Huxley 1972)

Given $0 \le \sigma \le 1$ and $T \ge 2$, define

$$N(\sigma, T) = \left|\{\rho = \beta + i\gamma : \zeta(\rho) = 0,\ \sigma \le \beta \le 1, |\gamma| \le T\}\right|.$$

There is an absolute constant $B > 0$ such that

$$N(\sigma, T) \ll T^{2.4(1-\sigma)}(\log T)^B.$$

Main ingredients

We need Stirling's formula of the form:

$$\Gamma(x+1) = \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \left(1 + O\left(\frac{1}{x}\right)\right).$$

The following is the zero density estimate by Huxley [H].

Lemma (Huxley 1972)

Given $0 \le \sigma \le 1$ and $T \ge 2$, define

$$N(\sigma, T) = \left|\{\rho = \beta + i\gamma : \zeta(\rho) = 0,\ \sigma \le \beta \le 1, |\gamma| \le T\}\right|.$$

There is an absolute constant $B > 0$ such that

$$N(\sigma, T) \ll T^{2.4(1-\sigma)}(\log T)^B.$$

Note that $N(\sigma, T) = 0$ for $\sigma > \frac{1}{2}$ if the Riemann Hypothesis is true.

Primes in almost short intervals

Denote by

$$L := L(X) = \frac{(\log X)^{1/3}}{(\log\log X)^{1/3}}.$$

Corollary

Let $X^{-5/6+\epsilon} \leq \delta \leq X^{-1/6}$. There is an absolute positive constant $c_0 := c_0(\epsilon) > 0$ such that for $x \in [X, 2X]$, $X \geq X_0(\epsilon)$

$$\pi(x+\delta x) - \pi(x) = \frac{\delta x}{\log x} + O\left(\delta x e^{-c_0 L}\right) \tag{1}$$

holds with an exceptional set $\mathcal{E}(X, \delta)$ of size at most $O(Xe^{-2c_0 L})$.

Toward multiple short intervals result

- We needed a similar result that applies to multiple short intervals.

Toward multiple short intervals result

• We needed a similar result that applies to multiple short intervals.
Problem 1. If we keep lengths of these short intervals the same, then δ must be allowed to vary with x.

Toward multiple short intervals result

• We needed a similar result that applies to multiple short intervals.
Problem 1. If we keep lengths of these short intervals the same, then δ must be allowed to vary with x.
Problem 2. If we keep δ not vary with x, then the lengths of the short intervals must vary.

Toward multiple short intervals result

• We needed a similar result that applies to multiple short intervals.
Problem 1. If we keep lengths of these short intervals the same, then δ must be allowed to vary with x.
Problem 2. If we keep δ not vary with x, then the lengths of the short intervals must vary.
We cannot have both the same length intervals and δ not varying with x.
So, if we try to fix 1, then 2 arises, and vice versa.

Whac-A-Mole

Whac-A-Mole


노래하는 두더지
1 위
2 위
점 수
!!순발력 테스트!!
300점 맞추면 더블게임
나의 순발력은 몇점?
판매.임대문의
Problem 2. Keep
the delta constant.
Problem 1. Keep
the same lengths.

Resolution: Just live with Problem 2.

Multiple almost all short intervals

Corollary

Let $X \geq X_0$, $\delta = X^{-1/2}e^{-c_0 L}$, $h = \lfloor 5e^{c_0 L}\log X \rfloor$, $x_0 := x_0(x) = \frac{x}{2} - \sqrt{X}\log X$, and $x_j := x_j(x) = (1+\delta)^j x_0$ for $j = 1, 2, \ldots, h$. Then there is a positive constant c_1 such that the set $\mathcal{E}(X)$ of all $x \in [X, 2X]$

$$\left| \pi(x_{j+1}) - \pi(x_j) - \frac{x_{j+1} - x_j}{\log x_j} \right| \geq (x_{j+1} - x_j)e^{-c_0 L}$$

for some $j = 0, 1, 2, \ldots, h-1$ satisfies $\mu(\mathcal{E}(X)) \ll Xe^{-c_1 L}$. Here, $\mu(A)$ is the Lebesgue measure of a set A.

Sketch of proof of almost short intervals result

Let $T = X^{5/6-\epsilon/2}$. Then

$$\psi(x+\delta x) - \psi(x) - \delta x = \sum_{|\Im(\rho)|\le T} \frac{(x+\delta x)^\rho - x^\rho}{\rho} + O(X^{1/6+\epsilon/2}(\log X)^2)$$
$$= \sum_{|\Im(\rho)|\le T} x^\rho w(\rho) + O(X^{1/6+2\epsilon/3}),$$

where $w(\rho) = \int_1^{1+\delta} u^{\rho-1} du$. Then we have

$$|\mathcal{E}(X,\delta)| \le \int_X^{2X} (\delta x)^{-2} e^{c_1 L} |\psi(x+\delta x) - \psi(x) - \delta x|^2 \, dx.$$

Sketch of proof of almost short intervals result

Let $T = X^{5/6-\epsilon/2}$. Then

$$\psi(x+\delta x) - \psi(x) - \delta x = \sum_{|\Im(\rho)|\le T} \frac{(x+\delta x)^\rho - x^\rho}{\rho} + O(X^{1/6+\epsilon/2}(\log X)^2)$$
$$= \sum_{|\Im(\rho)|\le T} x^\rho w(\rho) + O(X^{1/6+2\epsilon/3}),$$

where $w(\rho) = \int_1^{1+\delta} u^{\rho-1} du$. Then we have

$$|\mathcal{E}(X,\delta)| \le \int_X^{2X} (\delta x)^{-2} e^{c_1 L} |\psi(x+\delta x) - \psi(x) - \delta x|^2 \, dx.$$

Then we find an upper bound for the integral. We need the inequality

$$\int_X^{2X} x^{\beta_1+\beta_2+i(\gamma_1-\gamma_2)} \, dx \ll \frac{X^{\beta_1+\beta_2+1}}{|\gamma_1-\gamma_2|+1},$$

We use

$$\sum_{|\gamma_2|\le T} \frac{1}{|\gamma_1-\gamma_2|+1} \ll (\log T)^2$$

for any choice of γ_1. Then

$$\begin{aligned}\sum_{|\gamma|\le T} X^{2\beta} &= -\int_0^{1-\theta(T)} X^{2\sigma}\, dN(\sigma, T)\\ &\ll N(0,T) + \int_0^{1-\theta(T)} X^{2\sigma} N(\sigma, T)\, d\sigma\\ &\ll T(\log X) + (\log X)^B \int_0^{1-\theta(T)} X^{2\sigma} T^{2.4(1-\sigma)}\, d\sigma\\ &\ll X^2 e^{-c_2 L}.\end{aligned}$$

Here, $c_2 > 0$ is a constant which may depend on ϵ. We take $c_1 = c_2/2$. Then the result for $\psi(x)$ follows with $c_0 = c_1/2$.

Vinogradov's zero-free region

Let

$$\theta(T) := \frac{b}{(\log T)^{2/3}(\log\log T)^{1/3}}. \tag{2}$$

The constant $b > 0$ in $\theta(T)$ is given by Vinogradov's zero-free region for the Riemann zeta function so that

$$\beta < 1 - \theta(T)$$

for any zeta-zeros counted in $N(\sigma, T)$. Note that we can take $b = 1/57.54$ by [F].

Vinogradov's zero-free region

Let

$$\theta(T) := \frac{b}{(\log T)^{2/3}(\log\log T)^{1/3}}. \qquad (2)$$

The constant $b > 0$ in $\theta(T)$ is given by Vinogradov's zero-free region for the Riemann zeta function so that

$$\beta < 1 - \theta(T)$$

for any zeta-zeros counted in $N(\sigma, T)$. Note that we can take $b = 1/57.54$ by [F].
Note that the Riemann Hypothesis states all zeta-zeros satisfy $\beta = \frac{1}{2}$.

Proof of Theorem 1-continued

Let $\mathcal{E}(X)$ be the exceptional set. We treat S_x first over the intervals I_j and J defined as

$$I_j: \ (x_j, x_{j+1}] = \left(\frac{x}{2} + g(x_j)\sqrt{x}, \frac{x}{2} + g(x_{j+1})\sqrt{x}\right] \text{ for } j = 0, 1, 2, \dots h{-}1, \text{ and}$$

$$J: \ [0, x] - \bigcup_{j \leq h} I_j,$$

and $|g(x_j)| \leq 6 \log X$.

Proof of Theorem 1-continued

Let $\mathcal{E}(X)$ be the exceptional set. We treat S_x first over the intervals I_j and J defined as

$$I_j: \ (x_j, x_{j+1}] = \left(\frac{x}{2} + g(x_j)\sqrt{x}, \frac{x}{2} + g(x_{j+1})\sqrt{x}\right] \text{ for } j = 0, 1, 2, \ldots h{-}1, \text{ and}$$

$$J: \ [0, x] - \bigcup_{j \le h} I_j,$$

and $|g(x_j)| \le 6 \log X$.

- The sum over I_j is

$$\frac{g(x_{j+1}) - g(x_j)}{\log(x/2)} \frac{2}{\sqrt{2\pi}} e^{-2(g(x_j))^2} \left(1 + O\left((\log X)e^{-c_0 L}\right)\right).$$

Proof of Theorem 1-continued

Let $\mathcal{E}(X)$ be the exceptional set. We treat S_x first over the intervals I_j and J defined as

$I_j:\ (x_j, x_{j+1}] = \left(\frac{x}{2} + g(x_j)\sqrt{x}, \frac{x}{2} + g(x_{j+1})\sqrt{x}\right]$ for $j = 0, 1, 2, \ldots h{-}1$, and

$$J:\ [0, x] - \bigcup_{j\le h} I_j,$$

and $|g(x_j)| \le 6\log X$.

- The sum over I_j is

$$\frac{g(x_{j+1}) - g(x_j)}{\log(x/2)}\frac{2}{\sqrt{2\pi}}e^{-2(g(x_j))^2}\left(1 + O\left((\log X)e^{-c_0 L}\right)\right).$$

- The sum over J is by Hoeffding's inequality,

$$\frac{1}{2^x}\sum_{p\in J}\binom{x}{p} \le 4e^{-2(\log X)^2}.$$

Proof of Theorem 1-completion

We now take the sum over $j \leq h$. Then we have

$$\left| \sum_{j \leq h} (g(x_{j+1}) - g(x_j)) \frac{2}{\sqrt{2\pi}} e^{-2(g(x_j))^2} - \int_{-\infty}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-2u^2} \, du \right| \ll e^{-c_0 L}. \tag{3}$$

Putting together the sum over I_j's, we obtain

$$\left| \frac{1}{2^x} \sum_{p \leq x} \binom{x}{p} - \frac{1}{\log(x/2)} \right| \ll (\log X) e^{-c_0 L}.$$

The proof of Theorem 1.1 now follows

Proof of Theorem 2

Let $x \in [X, 2X]$, $h = \log X$, $c = 1/\log X$. Let I_j's and J defined by

$$I_j : \left(\frac{x}{2} + cj\sqrt{x}, \frac{x}{2} + c(j+1)\sqrt{x}\right] \text{ for } 0 \le |j| \le \frac{h}{c}, \text{ and}$$

$$J : [0, x] - \bigcup_{0 \le |j| \le \frac{h}{c}} I_j.$$

We apply Brun-Titchmarsh theorem [MV, Corollary 3.4] on each I_j.

$$\pi\left(\frac{x}{2} + c(j+1)\sqrt{x}\right) - \pi\left(\frac{x}{2} + cj\sqrt{x}\right) \le \frac{2c\sqrt{x}}{\log(c\sqrt{x})}\left(1 + O\left(\frac{1}{\log x}\right)\right)$$

$$= \frac{4c\sqrt{x}}{\log x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right).$$

Then the upper bound of the sum over I_j is given by

$$\frac{1}{2^x}\sum_{p\in I_j}\binom{x}{p} \leq \frac{4c\sqrt{x}}{\log x}\left(1+O\left(\frac{\log\log x}{\log x}\right)\right)\frac{2}{\sqrt{2\pi x}}e^{-2(cr_j)^2}, \qquad (4)$$

where $e^{-2(cr_j)^2} = \max_{x\in[cj,c(j+1)]} e^{-2x^2}$. Summing over $|j| \leq h/c$ and applying (13), we obtain

$$\frac{S_x}{2^x} - \frac{1}{2^x}\sum_{p\in J}\binom{x}{p} \leq \frac{1}{\log x}\left(4+O\left(\frac{\log\log x}{\log x}\right)\right) \qquad (5)$$

The sum over J is by Hoeffding's inequality,

$$\ll \frac{1}{(\log X)^2}.$$

The result follows.

Proof of Theorem 3

Let us first assume $0 < \alpha = \liminf_{n\to\infty} \frac{S_n \log n}{2^n}$. Then for sufficiently large n,

$$\frac{11\alpha}{12} \leq \frac{S_n \log n}{2^n}.$$

Let $c = 1/\log n$ and $h = \log n$. Then by Hoeffding's inequality,

$$\mathbf{P}\left(T_n \in \mathcal{P}, \left|T_n - \frac{n}{2}\right| \geq h\sqrt{n}\right) \leq \mathbf{P}\left(\left|T_n - \frac{n}{2}\right| \geq h\sqrt{n}\right) \leq 2e^{-2(\log n)^2}.$$

We use the subintervals I_j and J for $|j| \leq c/h$ as follows.

$$I_j: \ \left(\frac{n}{2} + cj\sqrt{n}, \frac{n}{2} + c(j+1)\sqrt{n}\right] \text{ for } 0 \leq |j| \leq \frac{h}{c}, \text{ and}$$

$$J: \ [0, n] - \bigcup_{0 \leq |j| \leq \frac{h}{c}} I_j.$$

Then

$$\sum_{p \in J} \binom{n}{p} \ll 2^n e^{-2(\log n)^2}.$$

Apply Brun-Titchmarsh inequality and choose $b_1 > 0$ so that the contribution of primes in the intervals I_j with $b_1 \leq c|j| \leq h$ is bounded by

$$\sum_{b_1 \leq c|j| \leq h} \sum_{p \in I_j} \binom{n}{p} \leq \frac{2^n}{\log n} \left(\int_{|t| \geq b_1} \frac{2}{\sqrt{2\pi}} e^{-2t^2} \; dt + O\left(\frac{1}{\log n}\right)\right) \leq \frac{2^n \alpha}{2 \log n}.$$

Then the contribution of primes in the interval $n/2 - b_1\sqrt{n} < p \leq n/2 + b_1\sqrt{n}$ is at least $\frac{2^n\alpha}{3\log n}$ for sufficiently large n. Thus,

$$\binom{n}{n/2}\left(\pi\left(\frac{n}{2} + b_1\sqrt{n}\right) - \pi\left(\frac{n}{2} - b_1\sqrt{n}\right)\right) \geq \frac{2^n\alpha}{3\log n}.$$

Then we have

$$\frac{2}{\sqrt{2\pi n}}\left(\pi\left(\frac{n}{2}+b_1\sqrt{n}\right)-\pi\left(\frac{n}{2}-b_1\sqrt{n}\right)\right)\geq\frac{\alpha}{3\log n}\left(1+O\left(\frac{h^3}{\sqrt{n}}\right)\right).$$

Now, this yields the lower bound for the number of primes in the short interval. There is $b_2>0$ such that for $n\geq N_0$,

$$\pi\left(\frac{n}{2}+b_1\sqrt{n}\right)-\pi\left(\frac{n}{2}-b_1\sqrt{n}\right)\geq\frac{b_2\sqrt{n}}{\log n}.$$

For the converse, assume that there are $b_1, b_2>0$ such that for $n\geq N_0$, the following holds.

$$\pi\left(\frac{n}{2}+b_1\sqrt{n}\right)-\pi\left(\frac{n}{2}-b_1\sqrt{n}\right)\geq\frac{b_2\sqrt{n}}{\log n}.$$

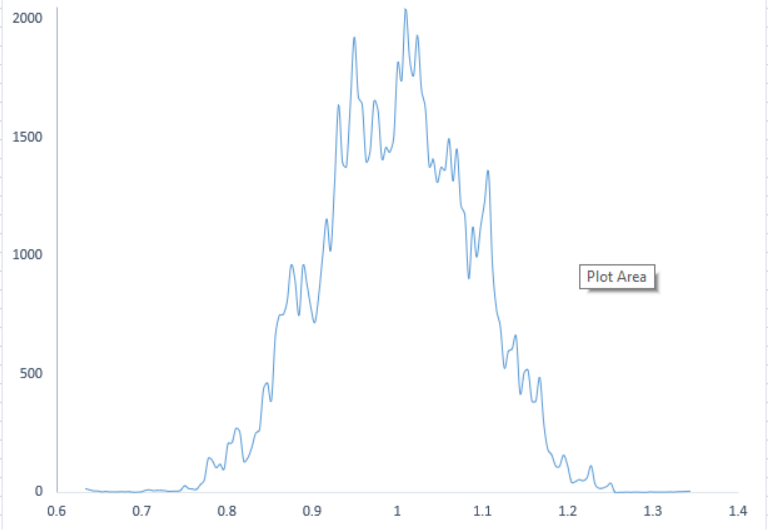
Then by Stirling's formula, there is an absolute constant $b_3>0$ such that,

$$S_n\geq\sum_{\left|\frac{n}{2}-p\right|\leq b_1\sqrt{n}}\binom{n}{p}\geq\binom{n}{\frac{n}{2}+b_1\sqrt{n}}\left(\pi\left(\frac{n}{2}+b_1\sqrt{n}\right)-\pi\left(\frac{n}{2}-b_1\sqrt{n}\right)\right)$$

$$\geq 2^n\frac{2}{\sqrt{2\pi n}}e^{-2b_1^2}\frac{b_2\sqrt{n}}{\log n}\left(1+O\left(\frac{h^3}{\sqrt{n}}\right)\right)\geq\frac{2^n b_3}{\log n}.$$

More numerical observations

• The distribution of $(S_n \log(n/2))/2^n$ for $n \leq 85000$ seems to be normal with mean 1 and standard deviation 0.09.



More results and speculations

• If we restrict the primes to arithmetic progression $a \mod q$ with $(a, q) = 1$, similar results hold with the factor $1/\phi(q)$.

More results and speculations

• If we restrict the primes to arithmetic progression $a \bmod q$ with $(a, q) = 1$, similar results hold with the factor $1/\phi(q)$.
• Currently, we do not have proof of the last observation. Proving any of these will require investigating the error term in the Theorem 1 asymptotic formula.

More results and speculations

• If we restrict the primes to arithmetic progression $a \bmod q$ with $(a, q) = 1$, similar results hold with the factor $1/\phi(q)$.
• Currently, we do not have proof of the last observation. Proving any of these will require investigating the error term in the Theorem 1 asymptotic formula.
• A corollary of Theorem (3) is that
$\alpha > 0$ implies a bound for prime gaps $g_n \ll \sqrt{p_n}$ which is stronger than Cramér's bound $g_n \ll \sqrt{p_n} \log p_n$ conditional on the Riemann Hypothesis (see [C]). To see this, for sufficiently large n, consider $m \in \mathbb{N}$ with $p_{n-1} \le m/2 - b_1\sqrt{m} < p_n$. Then we have $p_{n+1} \le m/2 + b_1\sqrt{m}$. Thus, $p_{n+1} - p_n < 2b_1\sqrt{m} \le C\sqrt{p_n}$.

More results and speculations

• If we restrict the primes to arithmetic progression $a \bmod q$ with $(a,q)=1$, similar results hold with the factor $1/\phi(q)$.
• Currently, we do not have proof of the last observation. Proving any of these will require investigating the error term in the Theorem 1 asymptotic formula.
• A corollary of Theorem (3) is that
$\alpha > 0$ implies a bound for prime gaps $g_n \ll \sqrt{p_n}$ which is stronger than Cramér's bound $g_n \ll \sqrt{p_n}\log p_n$ conditional on the Riemann Hypothesis (see [C]). To see this, for sufficiently large n, consider $m \in \mathbb{N}$ with $p_{n-1} \le m/2 - b_1\sqrt{m} < p_n$. Then we have $p_{n+1} \le m/2 + b_1\sqrt{m}$. Thus, $p_{n+1} - p_n < 2b_1\sqrt{m} \le C\sqrt{p_n}$.
• If we take the sequence $\lfloor ak^2 \rfloor$, $a>0$, $k=1,2,\ldots$ instead of primes, we observe an oscillating main term which resembles a solution to a heat equation.

References

H. Cramér, *On the distribution of primes*, Proc. Camb. Phil. Soc. 20, 272-280(1920).

K. Ford, *Zero Free Regions for the Riemann Zeta Function*, Number Theory for the Millenium (Urbana, IL, 2000) (M. A. Bennett, B. C. Berndt, N. Boston, H. G. Diamond, A. J. Hildebrand and W. Phillip, eds), vol. II (2002), 25-56.

M. N. Huxley, *On the Difference between Consecutive Primes*, Inventiones Math, 15, pp164–170(1972).

A. Kumchev, *The Distribution of Prime Numbers*, manuscript, 2005, URL: https://tigerweb.towson.edu/akumchev/distributionofprimesnotes.pdf

H. Montgomery, R. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge University Press 2007

N. K. Sinha, *What is the sum of the binomial coefficients* $\binom{n}{p}$ *over prime numbers?*, URL: https://math.stackexchange.com/q/2930593,