Is Goldbach's Conjecture Likely True?

Michael Emmerich
Faculty of Information Technology, University of Jyväskylä

8 August 2026

1 The question

Goldbach's conjecture says that every even integer greater than 2 is a sum of two primes. It is either true or false; there is no literal probability attached to its truth value. So what could it mean to say that the conjecture is "likely" to be true?

A useful interpretation is this: if we compare the primes with random or pseudo-random sets having similar density, how exceptional would a failure of the two-sum property be? From this point of view the evidence is striking. Random sets much sparser than the primes already tend to represent every sufficiently large even integer as a sum of two selected elements.

The primes therefore do not appear to be barely numerous enough. Their density lies far above the natural random threshold, and the usual Hardy–Littlewood heuristic predicts not one precarious representation of a large even integer N, but on the order of $N/(\log N)^2$ representations. Goldbach's conjecture has also been verified computationally to extremely large bounds [7].

The real question is therefore different. Can the deterministic arithmetic structure of the primes produce an exceptional even integer for which all of these possible representations fail simultaneously? The argument below separates these two issues: abundance is easy to explain probabilistically; ruling out a perfectly coordinated arithmetic failure is the difficult part.

2 A simple random model

Consider only odd positive integers. Include each odd integer n independently in a random set A with probability q_n.

For an even integer N, let $R_A(N)$ be the number of unordered representations

$$N = a + b, \qquad a, b \in A, \qquad a < b.$$

For fixed N, the possible pairs $\{a, N-a\}$ with $a < N/2$ are disjoint. Hence their selection events are independent. The expected number of representations is

$$m_N = \mathbb{E}R_A(N) = \sum_{\substack{3 \le a < N/2 \\ a \text{ odd}}} q_a q_{N-a}.$$

The probability that no pair is selected satisfies

$$\mathbb{P}(N \notin A + A) \le e^{-m_N}.$$

This is the essential density estimate. If m_N is eventually larger than a fixed multiple of $\log N$, the probabilities of missing successive even integers decrease fast enough that their sum is finite.

3 Borel–Cantelli and Chernoff in plain terms

The first Borel–Cantelli lemma is a bookkeeping principle for rare failures. If F_N is the event that N is missed and

$$\sum_N \mathbb{P}(F_N) < \infty,$$

then with probability 1 only finitely many of the events F_N occur. A summable bound for one integer at a time therefore becomes a statement about all sufficiently large integers at once. This direction does not require the events F_N themselves to be independent.

A Chernoff bound expresses a different idea. If X is the sum of many independent yes-or-no trials and $\mu = \mathbb{E}X$, then large relative deviations from μ have exponentially small probability. Schematically,

$$\mathbb{P}(|X - \mu| > \delta\mu) \le 2e^{-c_\delta \mu}.$$

Thus Chernoff bounds say that large independent sums stay close to their expected size, while Borel–Cantelli says that a summable sequence of exceptional probabilities can produce only finitely many exceptions almost surely. Readable introductions are given in [8] and [9].

4 The critical density

Near a large integer N, suppose the selection probability is roughly q_N. There are on the order of N candidate pairs and each succeeds with probability about q_N^2. Hence

$$m_N \asymp N q_N^2.$$

The transition relevant to eventual coverage occurs when m_N reaches the logarithmic scale:

$$N q_N^2 \asymp \log N.$$

Thus the natural critical density is

$$q_N \asymp \sqrt{\frac{\log N}{N}}.$$

This scale is classical in the theory of random additive bases [1,2]. The point here is not a new threshold theorem, but the comparison with the primes.

In the odd-integer model $q_n = c\sqrt{\log n/n}$, the elementary Borel–Cantelli calculation gives the sufficient boundary

$$c > \frac{2}{\sqrt{\pi}} \approx 1.128.$$

5 Prime density is much larger

Among odd integers near n, the local density of primes is approximately

$$q_n^{\text{prime}} \sim \frac{2}{\log n}.$$

Compared with the critical random scale,

$$\frac{2/\log n}{\sqrt{\log n/n}} \asymp \frac{\sqrt{n}}{(\log n)^{3/2}} \longrightarrow \infty.$$

At prime-like density the expected number of representations of a large even integer is of order

$$\frac{N}{(\log N)^2},$$

which is much larger than the logarithmic representation count near the threshold.

6 A three-way numerical comparison

Figure 1 compares three objects: the actual primes, one near-critical random set with

$$q_n = 1.45\sqrt{\frac{\log n}{n}},$$

and one independent prime-density random set with

$$q_n = \min\left(1, \frac{2}{\log n}\right).$$

The near-critical realization is only moderately above the elementary Borel–Cantelli boundary $2/\sqrt{\pi} \approx 1.128$. In 400 repeated samples, 98.5% represented every even

integer from 5,000 through 200,000. This is a finite-range numerical experiment, not an asymptotic proof.

For the particular plotted realizations, over the same even integers, the near-critical random set had median representation count 15, mean 14.8, and minimum 1. The prime-density random set had median 928, mean 912.8, and minimum 77. The actual primes had median 814, mean 908.2, and minimum 50.

The prime-density random sample and the actual primes are on the same broad abundance scale, while the near-critical sample is much lower, on the logarithmic scale.

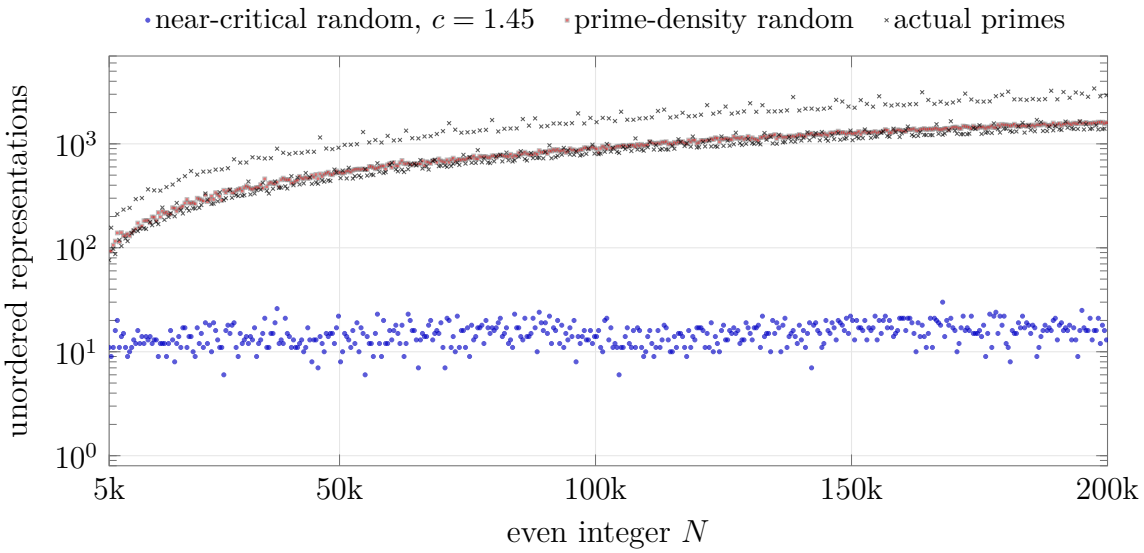

Figure 1: Representation counts in three models. The near-critical sample is plotted in blue. The prime-density random sample lies on the same broad scale as the actual primes but lacks the modular band structure visible in the prime counts. The plot uses every 400th even target for legibility; summary statistics in the text use every even integer from 5,000 to 200,000.

7 Why the prime counts form bands

The two dominant bands in the prime data are mainly a congruence effect modulo 3. The Hardy–Littlewood heuristic predicts a representation count containing the arithmetic factor

$$\prod_{\substack{p \mid N \\ p>2}} \frac{p-1}{p-2}.$$

If $3 \mid N$, the factor contributed by $p = 3$ is

$$\frac{3-1}{3-2} = 2.$$

Thus even integers divisible by 3 are expected, other things being comparable, to have roughly twice as many prime representations as nearby even integers not divisible by 3. Divisibility by 5, 7, and other small primes creates finer splittings inside the main bands.

The prime-density random sample has approximately the same abundance as the primes, but it does not know about divisibility by 3, 5, 7, and so on. Its representation counts are therefore much more homogeneous. Density explains the broad scale; arithmetic correlations explain the fine structure.

8 A much sparser subset of the primes

There is a useful deterministic comparison which shows that even full prime density is not necessary for strong additive behavior.

Consider the primes which can be written in the form

$$p = x^2 + y^2 + 1.$$

Iwaniec's work on primes represented by quadratic forms implies that the number of such primes up to x has order

$$\frac{x}{(\log x)^{3/2}}.$$

Since the total number of primes up to x has order $x/\log x$, this subclass has relative density of order

$$\frac{1}{\sqrt{\log x}} \longrightarrow 0$$

inside the primes. Thus it is genuinely sparse even when measured relative to the primes themselves.

Nevertheless, Teräväinen proved that almost every even integer satisfying the necessary local congruence conditions is a sum of two primes of the form $x^2 + y^2 + 1$ [6]. He also proved the corresponding ternary statement: every sufficiently large odd integer is a sum of three such primes.

This example sharpens the point of the density comparison. A set can have zero relative density inside the primes and still display very strong additive coverage, provided it is sufficiently well distributed and the local obstructions are respected.

There is also a useful comparison with the random threshold. The counting scale

$$\frac{x}{(\log x)^{3/2}}$$

is still much larger than the critical random counting scale

$$\sqrt{x \log x},$$

because

$$\frac{x/(\log x)^{3/2}}{\sqrt{x\log x}} = \frac{\sqrt{x}}{(\log x)^2} \longrightarrow \infty.$$

Thus Teräväinen's set is sparse relative to the primes, but it is not close to the random additive-basis threshold. The lesson is more subtle: density provides room for representations, while local arithmetic distribution determines whether that room can actually be used.

9 Why this does not prove Goldbach's conjecture

The primes are not independent random variables. They form one fixed deterministic sequence. A counterexample to Goldbach's conjecture would require that, for some even N, every prime $p < N/2$ has $N - p$ composite.

Informally, the primes would have to "conspire" so that all potential pairs fail at the same target. This is not a shortage-of-primes phenomenon. It is a statement about correlations.

A large prime gap by itself is not enough. One gap removes only a local collection of candidate pairs. A Goldbach counterexample requires the global reflected avoidance condition

$$p \text{ prime} \quad \Longrightarrow \quad N - p \text{ composite}$$

for every relevant prime p.

Density alone is also insufficient for deterministic subsets of the primes. Alsetri and Shao show that suitable local density hypotheses give an almost-all binary Goldbach theorem, while high global density alone can still fail badly [4]. Abundance and additive structure are different matters.

10 A logical asymmetry

Goldbach's conjecture is a universal assertion whose individual cases are finitely decidable. If it is false, there is a definite even integer N_* which is a counterexample, and that failure can be checked by finite computation. Falsity therefore has a finite certificate.

If the conjecture is true, there is no last integer to check. A proof has to control all even integers at once. In principle a true universal arithmetic statement can even be unprovable in a chosen formal theory, although there is no evidence that this is what happens for Goldbach's conjecture.

11 Could the reason simply be statistical?

There need not be a simple distinguished mechanism assigning a special prime pair to every even integer. It is conceivable that Goldbach's conjecture is true for a more diffuse reason: after the necessary congruence restrictions are taken into account, the primes may be sufficiently dense and sufficiently pseudorandom that the many possible representations never all disappear at once.

In that sense the truth could be statistical in character. At prime density one expects on the order of

$$\frac{N}{(\log N)^2}$$

candidate representations. A failure would require the arithmetic correlations of the primes to suppress all of them simultaneously.

But a statistical-looking reason for truth does not imply that there is no finite proof. A finite proof need not identify a canonical representation of each N. It could instead prove a uniform statement such as $R(N) > 0$ for all sufficiently large N by establishing enough deterministic pseudorandomness of the primes. Analytic number theory often turns statistical regularity into finite proofs.

Thus one should keep two questions separate:

$$\text{no simple structural explanation} \quad \not\Longrightarrow \quad \text{no finite proof.}$$

Unprovability is a different, theory-relative possibility. Goldbach's conjecture could in principle be true but unprovable in a specified formal system such as Peano arithmetic. That possibility comes from mathematical logic, not from the apparent randomness of the primes. There is presently no evidence that Goldbach's conjecture is independent of the usual formal systems.

The contrast is nevertheless attractive: the truth may be statistical in character, while any proof must be deterministic.

Related work. A closely related probabilistic perspective was presented by Stuart Armstrong in a 2020 LessWrong essay, "The Goldbach conjecture is probably correct; so was Fermat's last theorem" [11]. Starting from the prime number theorem, Armstrong gives a random-sums heuristic in which the probability that a large target N receives no representation is estimated on a scale comparable to

$$\exp\left(-\frac{N}{2(\log N)^2}\right).$$

He also emphasizes the same interpretive caution used here: this is not a literal probability that a deterministic proposition is true, but a typicality statement about prime-like sequences. The present article develops a different part of that idea by putting the comparison into an explicit independent random-set model and deriving the additive-basis threshold

$$q_N \asymp \sqrt{\frac{\log N}{N}}.$$

A complementary rigorous landmark is Helfgott's proof of the ternary, or weak, Goldbach conjecture [12]: every odd integer greater than 5 is a sum of three primes. His argument uses the circle method together with the large sieve and explicit estimates for exponential sums over primes. The ternary theorem does not imply the binary Goldbach conjecture considered here. It does, however, illustrate the central distinction in this article: heuristic abundance is only the beginning; a proof must obtain enough deterministic control over the arithmetic distribution of the primes to make the desired additive statement hold uniformly.

References

1. P. Erdős, "On a problem of Sidon in additive number theory," *Acta Scientiarum Mathematicarum (Szeged)* 15 (1954), 255–259.

2. P. Erdős and A. Rényi, "Additive properties of random sequences of positive integers," *Acta Arithmetica* 6 (1960), 83–110.

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 (1923), 1–70.

4. A. Alsetri and X. Shao, "Density versions of the binary Goldbach problem," *Acta Arithmetica* 218 (2025), 285–295.

5. H. Iwaniec, "Primes of the type $\phi(x, y) + A$ where ϕ is a quadratic form," *Acta Arithmetica* 21 (1972), 203–234. DOI: 10.4064/aa-21-1-203-234.

6. J. Teräväinen, "The Goldbach problem for primes that are sums of two squares plus one," *Mathematika* 64 (2018), 20–70. DOI: 10.1112/S0025579317000341.

7. T. Oliveira e Silva, S. Herzog, and S. Pardi, "Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$," *Mathematics of Computation* 83 (2014), 2033–2060.

8. M. Mitzenmacher and E. Upfal, *Probability and Computing*, 2nd ed., Cambridge University Press, 2017, Chapter 4.

9. G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed., Oxford University Press, 2001.

10. R. Kaye, *Models of Peano Arithmetic*, Oxford University Press, 1991.

11. S. Armstrong, "The Goldbach conjecture is probably correct; so was Fermat's last theorem," *LessWrong*, 14 July 2020.

12. H. A. Helfgott, "The ternary Goldbach conjecture is true," arXiv:1312.7748, 2013. See also H. A. Helfgott, *The ternary Goldbach problem*, arXiv:1501.05438, 2015.

A Technical appendix: the random model behind the article

These appendices supply the probability calculations that were only summarized in the blog post. They are deliberately about the *random model*. None of them turns the random model into a proof of Goldbach's conjecture for the actual primes.

For asymptotic statements it is convenient to make the selection probabilities legitimate for every odd integer by writing

$$q_n = \min\left(1, c\sqrt{\frac{\log n}{n}}\right).$$

The truncation at 1 changes only finitely many small values and therefore has no effect on any large-n conclusion below.

A.1 A fixed even target

Fix an even integer N. For each odd a with $3 \le a < N/2$, define

$$I_a = \mathbf{1}_{\{a\in A\}}\,\mathbf{1}_{\{N-a\in A\}}.$$

Then $I_a = 1$ exactly when the pair $\{a, N-a\}$ gives a representation of N. Hence

$$R_A(N) = \sum_{\substack{3\le a<N/2\\ a \text{ odd}}} I_a.$$

Because the pairs $\{a, N-a\}$ are disjoint as a ranges below $N/2$, the variables I_a are independent. Moreover

$$\mathbb{P}(I_a = 1) = q_a q_{N-a}.$$

Thus $R_A(N)$ is a sum of independent, not necessarily identically distributed, Bernoulli random variables and

$$m_N := \mathbb{E}R_A(N) = \sum_{\substack{3\le a<N/2\\ a \text{ odd}}} q_a q_{N-a}.$$

This observation is what makes the elementary product estimate and the Chernoff bounds available.

A.2 The exact zero-representation estimate

Independence gives

$$\begin{aligned}\mathbb{P}(R_A(N) = 0) &= \prod_{\substack{3\le a<N/2\\ a \text{ odd}}} \mathbb{P}(I_a = 0)\\ &= \prod_{\substack{3\le a<N/2\\ a \text{ odd}}} \bigl(1 - q_a q_{N-a}\bigr).\end{aligned}$$

For $x \geq 0$ we have $1 - x \leq e^{-x}$ whenever $0 \leq x \leq 1$. One proof is to set

$$f(x) = e^{-x} - (1 - x).$$

Then $f(0) = 0$ and $f'(x) = 1 - e^{-x} \geq 0$ for $x \geq 0$, so $f(x) \geq 0$. Applying this factor by factor yields

$$\begin{aligned}\mathbb{P}(R_A(N) = 0) &\leq \prod_a e^{-q_a q_{N-a}} \\ &= \exp\left(-\sum_a q_a q_{N-a}\right) \\ &= e^{-m_N}.\end{aligned}$$

This is the bound used in the main text. Notice that for the event $R_A(N) = 0$ it is stronger than the commonly quoted quadratic lower-tail Chernoff bound; Appendix E explains the relationship.

B Deriving the critical density and the constant $2/\sqrt{\pi}$

There are two related statements in the phrase "critical density." First comes the scale

$$q_N \asymp \sqrt{\frac{\log N}{N}}.$$

Second, for the particular independent odd-integer model used in the article, the elementary Borel–Cantelli argument produces the explicit sufficient coefficient

$$c > \frac{2}{\sqrt{\pi}}.$$

We derive both carefully.

B.1 Why the scale is $\sqrt{\log N/N}$

Suppose, just for this first-order calculation, that all relevant selection probabilities near a large target N have roughly the same size q_N. There are a constant multiple of N unordered odd pairs $a + (N - a) = N$. Each pair is selected with probability roughly q_N^2. Therefore

$$m_N \asymp N q_N^2.$$

A missed target has probability at most e^{-m_N}. To make the failure probabilities comparable with a summable power $N^{-1-\eta}$, one wants m_N on the order of $\log N$. Equating the two scales gives

$$N q_N^2 \asymp \log N,$$

and hence

$$\boxed{q_N \asymp \sqrt{\frac{\log N}{N}}.}$$

This argument determines the scale but not the leading constant. The leading constant requires evaluating the sum defining m_N.

B.2 Asymptotics of the mean representation count

Take

$$q_n = c\sqrt{\frac{\log n}{n}}$$

for large odd n. We claim that

$$m_N \sim \frac{\pi c^2}{4}\log N \qquad (N \to \infty,\ N \text{ even}). \tag{1}$$

To see where the constant comes from, fix $0 < \varepsilon < 1/2$ and first restrict to $\varepsilon N \le a < N/2$. On this range

$$\frac{\log a}{\log N} \to 1, \qquad \frac{\log(N-a)}{\log N} \to 1$$

uniformly as $N \to \infty$. Hence the central part of the sum is

$$\begin{aligned} m_N^{(\varepsilon)} &= (1+o(1))c^2 \log N \sum_{\substack{\varepsilon N \le a < N/2 \\ a \text{ odd}}} \frac{1}{\sqrt{a(N-a)}} \\ &= (1+o(1))\frac{c^2 \log N}{N} \sum_{\substack{\varepsilon N \le a < N/2 \\ a \text{ odd}}} \frac{1}{\sqrt{(a/N)(1-a/N)}}. \end{aligned}$$

The odd integers have mesh size $2/N$ after scaling by N. Therefore

$$\frac{1}{N} \sum_{\substack{\varepsilon N \le a < N/2 \\ a \text{ odd}}} f(a/N) \longrightarrow \frac{1}{2}\int_{\varepsilon}^{1/2} f(t)\,dt,$$

with $f(t) = 1/\sqrt{t(1-t)}$. It follows that

$$m_N^{(\varepsilon)} \sim \frac{c^2 \log N}{2}\int_{\varepsilon}^{1/2} \frac{dt}{\sqrt{t(1-t)}}.$$

The integral is elementary. With $t = \sin^2\theta$,

$$\frac{dt}{\sqrt{t(1-t)}} = 2\,d\theta,$$

so

$$\int_0^{1/2} \frac{dt}{\sqrt{t(1-t)}} = 2\arcsin\sqrt{\frac{1}{2}} = \frac{\pi}{2}.$$

Consequently, after letting $\varepsilon \downarrow 0$, the central contribution approaches

$$\frac{c^2 \log N}{2} \cdot \frac{\pi}{2} = \frac{\pi c^2}{4}\log N.$$

It remains to check that the omitted endpoint $a < \varepsilon N$ does not alter the leading constant. Since $N - a \geq N/2$ there, its contribution is bounded by a constant multiple of

$$c^2 \sqrt{\frac{\log N}{N}} \sum_{a \leq \varepsilon N} \sqrt{\frac{\log a}{a}}.$$

An integral comparison gives

$$\sum_{a \leq y} \sqrt{\frac{\log a}{a}} = O\Big(\sqrt{y \log y}\Big),$$

so the endpoint contributes only

$$O\Big(c^2 \sqrt{\varepsilon}\, \log N\Big).$$

Letting $\varepsilon \downarrow 0$ proves (1).

B.3 From the mean to the Borel–Cantelli boundary

By Appendix A,

$$\mathbb{P}(R_A(N) = 0) \leq e^{-m_N}.$$

Using (1),

$$\mathbb{P}(R_A(N) = 0) \leq N^{-\pi c^2/4 + o(1)}.$$

If

$$\frac{\pi c^2}{4} > 1,$$

then one can choose a number $\eta > 0$ such that, for all sufficiently large even N,

$$\mathbb{P}(R_A(N) = 0) \leq N^{-1-\eta}.$$

The series $\sum_N N^{-1-\eta}$ converges, so the first Borel–Cantelli lemma implies that only finitely many even targets are missed almost surely. Solving the coefficient inequality gives

$$\boxed{c > \frac{2}{\sqrt{\pi}} \approx 1.128379.}$$

This constant should be interpreted precisely. It is the boundary delivered by this elementary expectation/product/Borel–Cantelli argument for this particular model. At $c = 2/\sqrt{\pi}$ the bound is of harmonic order $1/N$ and is no longer summable. For $c < 2/\sqrt{\pi}$ this calculation stops proving eventual coverage; it does *not* by itself prove that infinitely many failures occur, nor is it being asserted here as the sharp threshold theorem for every formulation of a random additive basis.

B.4 The corresponding counting scale

The expected number of selected odd integers up to x is

$$\mathbb{E}|A \cap [1,x]| = \sum_{\substack{n \le x \\ n \text{ odd}}} q_n.$$

Using the same odd-lattice factor $1/2$ and the integral estimate

$$\int_2^x \sqrt{\frac{\log t}{t}}\, dt \sim 2\sqrt{x \log x},$$

one obtains

$$\mathbb{E}|A \cap [1,x]| \sim c\sqrt{x \log x}.$$

Thus the density statement $q_n \asymp \sqrt{\log n / n}$ and the counting statement $|A \cap [1,x]| \asymp \sqrt{x \log x}$ are two versions of the same critical scale.

C Union bounds and the first Borel–Cantelli lemma

C.1 The union bound

For events $E_1, \dots, E_M$,

$$\mathbf{1}_{\cup_{j=1}^M E_j} \le \sum_{j=1}^M \mathbf{1}_{E_j}.$$

Taking expectations gives

$$\boxed{\mathbb{P}\left(\bigcup_{j=1}^M E_j\right) \le \sum_{j=1}^M \mathbb{P}(E_j).}$$

No independence is required.

C.2 First Borel–Cantelli lemma, with proof

Let $F_1, F_2, \dots$ be any events and suppose

$$\sum_{n=1}^{\infty} \mathbb{P}(F_n) < \infty.$$

The event that infinitely many F_n occur is

$$\limsup_{n\to\infty} F_n = \bigcap_{m=1}^{\infty} \bigcup_{n \ge m} F_n.$$

For every m, the union bound gives

$$\mathbb{P}\left(\bigcup_{n \ge m} F_n\right) \le \sum_{n \ge m} \mathbb{P}(F_n).$$

Because the numerical series converges, its tail tends to zero. The events $\bigcup_{n\geq m} F_n$ decrease with m, so continuity of probability from above yields

$$\begin{aligned}\mathbb{P}(F_n \text{ infinitely often}) &= \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} F_n\right) \\ &\leq \lim_{m\to\infty} \sum_{n\geq m} \mathbb{P}(F_n) \\ &= 0.\end{aligned}$$

Therefore, with probability 1, only finitely many of the F_n occur. This proof also makes clear why independence between the failure events for different target integers is unnecessary.

D A didactic derivation of Markov's inequality

Markov's inequality is the elementary engine behind the Chernoff method. It converts information about an average into an upper bound on a tail probability.

D.1 Statement

Let Y be a nonnegative random variable with finite expectation, and let $a > 0$. Then

$$\boxed{\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}Y}{a}.} \tag{2}$$

D.2 Proof by an indicator

Pointwise, for every outcome,

$$Y \geq a\,\mathbf{1}_{\{Y\geq a\}}.$$

Indeed, if $Y < a$, the right-hand side is 0 and the inequality uses $Y \geq 0$; if $Y \geq a$, the right-hand side equals a and the inequality is again true. Taking expectations preserves the inequality:

$$\mathbb{E}Y \geq a\,\mathbb{E}\mathbf{1}_{\{Y\geq a\}}.$$

But the expectation of an indicator is its event probability, so

$$\mathbb{E}\mathbf{1}_{\{Y\geq a\}} = \mathbb{P}(Y \geq a).$$

Hence

$$\mathbb{E}Y \geq a\,\mathbb{P}(Y \geq a),$$

and division by $a > 0$ proves (2).

D.3 Why nonnegativity is necessary

The pointwise comparison $Y \geq a\mathbf{1}_{\{Y \geq a\}}$ can fail on the complement of the event if Y is allowed to be negative. Markov's inequality is therefore fundamentally a statement about nonnegative random variables. Chernoff's method gets around this limitation for an arbitrary real random variable X by applying Markov to e^{tX}, which is always positive.

D.4 Exponential Markov inequality

Let X be any random variable for which $\mathbb{E}e^{tX}$ is finite, and let $t > 0$. Since the exponential is increasing,

$$\{X \geq a\} = \{e^{tX} \geq e^{ta}\}.$$

Applying Markov to $Y = e^{tX}$ gives

$$\boxed{\mathbb{P}(X \geq a) \leq e^{-ta}\,\mathbb{E}e^{tX}.} \tag{3}$$

The right-hand side is valid for every $t > 0$, so we may choose the value of t that makes it as small as possible. That optimization step is the essence of a Chernoff bound.

For a lower tail, take $t > 0$ and apply Markov to e^{-tX}:

$$\mathbb{P}(X \leq a) \leq e^{ta}\,\mathbb{E}e^{-tX}. \tag{4}$$

E A didactic derivation of Chernoff bounds

Let

$$X = \sum_{i=1}^{n} X_i,$$

where X_i are independent Bernoulli random variables with

$$\mathbb{P}(X_i = 1) = p_i, \qquad \mathbb{P}(X_i = 0) = 1 - p_i.$$

They need not have the same success probability. Put

$$\mu = \mathbb{E}X = \sum_{i=1}^{n} p_i.$$

E.1 Step 1: the Bernoulli moment generating function

For one Bernoulli variable,

$$\begin{aligned} \mathbb{E}e^{tX_i} &= (1 - p_i)e^0 + p_i e^t \\ &= 1 + p_i(e^t - 1). \end{aligned}$$

Using $1 + u \leq e^u$ for every real u gives

$$\mathbb{E}e^{tX_i} \leq \exp\Big(p_i(e^t - 1)\Big).$$

E.2 Step 2: independence factorizes the MGF

Because the X_i are independent,

$$\begin{aligned}\mathbb{E}e^{tX} &= \mathbb{E}e^{t\sum_i X_i} \\ &= \mathbb{E}\prod_i e^{tX_i} \\ &= \prod_i \mathbb{E}e^{tX_i} \\ &\le \prod_i \exp\Big(p_i(e^t-1)\Big) \\ &= \exp\Big(\mu(e^t-1)\Big).\end{aligned}$$

Thus

$$\boxed{\mathbb{E}e^{tX} \le e^{\mu(e^t-1)}.} \tag{5}$$

E.3 Upper tail: exact multiplicative form

Let $\delta > 0$. From exponential Markov and (5), for every $t > 0$,

$$\mathbb{P}\Big(X \ge (1+\delta)\mu\Big) \le \exp\Big(\mu(e^t-1) - t(1+\delta)\mu\Big).$$

Differentiate the exponent with respect to t:

$$\mu e^t - (1+\delta)\mu.$$

The minimizing value therefore satisfies $e^t = 1+\delta$, i.e.

$$t = \log(1+\delta).$$

Substitution yields

$$\begin{aligned}\mathbb{P}\Big(X \ge (1+\delta)\mu\Big) &\le \exp\Big(-\mu\Big[(1+\delta)\log(1+\delta) - \delta\Big]\Big) \\ &= \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.\end{aligned}$$

Hence

$$\boxed{\mathbb{P}\Big(X \ge (1+\delta)\mu\Big) \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.} \tag{6}$$

E.4 Upper tail: the familiar quadratic form

For $\delta \ge 0$,

$$\log(1+\delta) \ge \frac{2\delta}{2+\delta}.$$

Indeed, the difference has value 0 at $\delta = 0$ and derivative

$$\frac{1}{1+\delta} - \frac{4}{(2+\delta)^2} = \frac{\delta^2}{(1+\delta)(2+\delta)^2} \ge 0.$$

Therefore

$$(1+\delta)\log(1+\delta) - \delta \geq (1+\delta)\frac{2\delta}{2+\delta} - \delta$$
$$= \frac{\delta^2}{2+\delta}.$$

So

$$\mathbb{P}\Big(X \geq (1+\delta)\mu\Big) \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right). \tag{7}$$

In particular, if $0 \leq \delta \leq 1$, then $2+\delta \leq 3$ and

$$\boxed{\mathbb{P}\Big(X \geq (1+\delta)\mu\Big) \leq e^{-\delta^2\mu/3}.} \tag{8}$$

E.5 Lower tail: exact multiplicative form

Let $0 < \delta < 1$. Apply (4) to the threshold $(1-\delta)\mu$. For $t > 0$,

$$\mathbb{E}e^{-tX} \leq \exp\Big(\mu(e^{-t}-1)\Big),$$

so

$$\mathbb{P}\Big(X \leq (1-\delta)\mu\Big) \leq \exp\Big(\mu(e^{-t}-1) + t(1-\delta)\mu\Big).$$

Differentiate the exponent:

$$-\mu e^{-t} + (1-\delta)\mu.$$

The minimizing value satisfies $e^{-t} = 1-\delta$, hence

$$t = -\log(1-\delta).$$

Substituting,

$$\boxed{\mathbb{P}\Big(X \leq (1-\delta)\mu\Big) \leq \exp\Big(-\mu\Big[\delta + (1-\delta)\log(1-\delta)\Big]\Big).} \tag{9}$$

E.6 Lower tail: the quadratic form

Define

$$h(\delta) = \delta + (1-\delta)\log(1-\delta) - \frac{\delta^2}{2}.$$

We have $h(0) = 0$, and

$$h'(\delta) = -\log(1-\delta) - \delta.$$

Since $-\log(1-\delta) \geq \delta$ for $0 \leq \delta < 1$, it follows that $h'(\delta) \geq 0$. Therefore

$$\delta + (1-\delta)\log(1-\delta) \geq \frac{\delta^2}{2}.$$

Combining this with (9) gives

$$\boxed{\mathbb{P}\Big(X \leq (1-\delta)\mu\Big) \leq e^{-\delta^2\mu/2}, \qquad 0 < \delta < 1.} \tag{10}$$

E.7 Two-sided concentration

For $0 < \delta \leq 1$, the union bound gives

$$\begin{aligned}\mathbb{P}(|X-\mu| \geq \delta\mu) &\leq \mathbb{P}(X \geq (1+\delta)\mu) + \mathbb{P}(X \leq (1-\delta)\mu) \\ &\leq e^{-\delta^2\mu/3} + e^{-\delta^2\mu/2} \\ &\leq 2e^{-\delta^2\mu/3}.\end{aligned}$$

Thus

$$\boxed{\mathbb{P}(|X-\mu| \geq \delta\mu) \leq 2e^{-\delta^2\mu/3}, \qquad 0 < \delta \leq 1.} \tag{11}$$

E.8 The special event $X = 0$

For independent Bernoulli variables there is an especially clean calculation:

$$\begin{aligned}\mathbb{P}(X=0) &= \prod_{i=1}^{n}(1-p_i) \\ &\leq \prod_{i=1}^{n} e^{-p_i} \\ &= e^{-\sum_i p_i} \\ &= e^{-\mu}.\end{aligned}$$

Hence

$$\boxed{\mathbb{P}(X=0) \leq e^{-\mu}.} \tag{12}$$

This is exactly the estimate used for a missed Goldbach target in the random model. It is also the endpoint of the exact lower-tail Chernoff rate: as $\delta \uparrow 1$,

$$\delta + (1-\delta)\log(1-\delta) \longrightarrow 1,$$

so (9) tends to $e^{-\mu}$.

F Finite-range reliability thresholds

The asymptotic threshold in Appendix B answers an "eventually for all targets" question. For a finite numerical experiment it is often more useful to ask a different question: how large should the expected number of representations be in order to make *any* failure among M targets unlikely?

Suppose $X_1, \ldots, X_M$ are representation counts for M targets, and suppose each has mean at least μ. The counts need not be independent across targets. If the fixed-target Bernoulli structure gives

$$\mathbb{P}(X_j = 0) \leq e^{-\mu},$$

then the union bound gives

$$\mathbb{P}(\text{at least one target is missed}) \leq Me^{-\mu}.$$

To make this at most a prescribed error probability ε, it is enough that

$$Me^{-\mu} \le \varepsilon.$$

Taking logarithms gives the exact zero-event reliability threshold

$$\boxed{\mu \ge \log \frac{M}{\varepsilon}.} \tag{13}$$

If one forgets the special product formula and uses only the quadratic lower-tail Chernoff estimate at the endpoint $\delta = 1$, one gets the weaker bound

$$\mathbb{P}(X_j = 0) \le e^{-\mu/2}.$$

Then it is sufficient that

$$\boxed{\mu \ge 2\log \frac{M}{\varepsilon}.} \tag{14}$$

The factor of two is the price of using the convenient quadratic approximation instead of the exact zero-event exponent.

These formulas explain a central theme of the blog post in a compact way: if a target has many roughly independent opportunities for representation, then the expected number of opportunities need grow only logarithmically with the number of targets one wants to protect simultaneously.

G How the appendices connect back to Goldbach

For one fixed even N in the random odd-integer model, everything above is rigorous: the candidate-pair indicators are independent, their mean is m_N, the missing probability is at most e^{-m_N}, and the full Chernoff theory applies to $R_A(N)$.

For different targets N, the corresponding representation counts share underlying membership variables and are generally dependent. This is harmless for the union bound and for the first Borel–Cantelli lemma, neither of which requires independence between targets.

The decisive limitation is elsewhere: primality is not generated by independent Bernoulli trials. The actual primes obey congruence restrictions and long-range arithmetic correlations. Consequently the random calculation is evidence about abundance and about what a pseudorandom set of comparable density would do. It is not a transfer theorem from the random model to the primes. A proof of Goldbach's conjecture would have to control the deterministic arithmetic correlations strongly enough to rule out a simultaneous failure of every candidate pair.