Erdos problem 205: de-formalization of Aristotle's proof

Nat Sothanaphan*

Jan 11, 2026

Introduction

Erdos problem 205 is stated in `https://www.erdosproblems.com/205`. In that thread, following Woett's suggestion, Leeham obtained a formal proof by Aristotle. Later, Terence Tao suggested that an improved asymptotics is possible; Boris Alexeev then formalized this improved result, also using Aristotle. The improved formalization is available at `https://github.com/plby/lean-proofs/blob/main/ErdosProblems/Erdos205.md`.

The author, using ChatGPT[1], de-formalized the improved formalization as this present writeup. Note that the files linked in the shared chat are different from this present writeup; the present writeup contains many additional improvements by the author. The author has checked everything manually.

Assumptions and definitions

Definition 1 (nth_prime). *For $n \in \mathbb{N}$, let* $\text{nth\_prime}(n)$ *denote the n-th prime number (with the convention that $n = 0$ gives 2, $n = 1$ gives 3, etc.).*

Lemma 1 (Prime number theorem input: `nth_prime_asymp`). *We assume as an axiom that, as $n \to \infty$,*

$$\text{nth\_prime}(n) \sim n \log n.$$

Definition 2 (`Omega`). *For $m \in \mathbb{N}$, let $\Omega(m)$ be the number of prime factors of m, counted with multiplicity. Equivalently, if $m = \prod_p p^{v_p(m)}$ is the prime factorization, then*

$$\Omega(m) = \sum_p v_p(m).$$

Definition 3 (`pntRate`). *Define the "PNT-scale rate"*

$$\text{pntRate}(n) := \sqrt{\frac{\log n}{\log \log n}},$$

interpreting $\log$ *and* $\sqrt{}$ *as the usual real functions (and ignoring the small-n edge cases, which the Lean file handles via total functions).*

*natsothanaphan@gmail.com

[1]https://chatgpt.com/share/69639adf-d56c-8010-bf7b-c17c35ce0955

Lemma 2 (`nth_prime_le_const_mul_log_eventually`). *From the asymptotic* $\text{nth\_prime}(n) \sim n \log n$*, one can extract an eventual upper bound: there exists a constant* $C > 0$ *and an* N *such that for all* $n \geq N$*,*

$$\text{nth\_prime}(n) \leq C\, n \log n.$$

Proof. Since $\text{nth\_prime}(n)/(n \log n) \to 1$, the ratio is eventually < 2. Taking $C = 2$ yields the claim. □

The construction

Fix an integer $E \geq 10$.

Definition 4 (`p_kj`). *For* $k, j \in \mathbb{N}$*, define*

$$p_{k,j} \;:=\; \text{nth\_prime}\big(kE + (j-1) + 2\big).$$

Remark 1. *The index shift "+2" means* $p_{k,j} \geq 5$*, so these primes are odd and not equal to* 3*. As* k *runs over* $0, \ldots, E-1$ *and* j *runs over* $1, \ldots, E$*, we obtain* E^2 *distinct primes.*

Definition 5 (`Q_k`). *For* $k \in \mathbb{N}$*, define*

$$Q_k \;:=\; \prod_{j=1}^{E} p_{k,j}.$$

Definition 6 (`M`). *Define the total modulus*

$$M \;:=\; 2^E \cdot 3 \cdot \prod_{k=0}^{E-1} Q_k.$$

Definition 7 (`is_solution`). *We say* $n \in \mathbb{N}$ *is a solution (for this* E*) if the following congruences hold:*

$$n \equiv 0 \pmod{2^E}, \qquad n \equiv 0 \pmod{3}, \qquad n \equiv 2^k \pmod{Q_k} \;\; \textit{for every } k = 0, 1, \ldots, E-1.$$

Elementary properties of the moduli

Lemma 3 (`p_kj_ge_5`). *For all* E, k, j*, we have* $p_{k,j} \geq 5$*.*

Proof. Trivial. □

Lemma 4 (`p_kj_injective`). *For fixed* E*, the map* $(k, j) \mapsto p_{k,j}$ *is injective on the index set*

$$0 \leq k \leq E-1, \quad 1 \leq j \leq E.$$

Proof. Trivial. □

Lemma 5 (`Q_k_props`). *For* $0 \leq k \leq E-1$*, the integer* Q_k *is squarefree and satisfies* $\Omega(Q_k) = E$*.*

Proof. Trivial. □

Lemma 6 (`moduli_properties`). *For $0 \le k \le E-1$:*

1. *Q_k is coprime to 2,*
2. *Q_k is coprime to 3,*
3. *if $k_1 \ne k_2$ then $\gcd(Q_{k_1}, Q_{k_2}) = 1$.*

Proof. Trivial. □

Chinese remainder theorem

The Lean development packages the congruences as a list of moduli and remainders: index 0 corresponds to $(2^E, 0)$, index 1 to $(3, 0)$, and index $k+2$ to $(Q_k, 2^k)$.

Lemma 7 (`moduli_pairwise_coprime`). *The list of moduli $(2^E, 3, Q_0, \ldots, Q_{E-1})$ is pairwise coprime.*

Proof. From Lemma `moduli_properties` and $\gcd(2^E, 3) = 1$. □

Definition 8 (`n_E`). *Let n_E be the (canonical) CRT solution modulo the product $2^E \cdot 3 \cdot \prod_{k<E} Q_k$ to the congruence system:*

$$n \equiv 0 \pmod{2^E}, \quad n \equiv 0 \pmod 3, \quad n \equiv 2^k \pmod{Q_k} \quad (0 \le k \le E-1).$$

Lemma 8 (`n_E_is_solution`). *The integer n_E is indeed a solution in the sense of* `is_solution`.

Proof. This verifies the CRT construction. □

Basic lemmas about Ω

Lemma 9 (`Omega_dvd_le`). *If $a \mid b$ and $b \ne 0$, then $\Omega(a) \le \Omega(b)$.*

Proof. Follows from definition. □

Lemma 10 (`Omega_pow_two`). *For all E, $\Omega(2^E) = E$.*

Proof. Follows from definition. □

Lemma 11 (`n_E_not_pow_two`). *For every k, $n_E \ne 2^k$.*

Proof. Follows from $n_E \equiv 0 \pmod 3$. □

The core Ω lower bound

Lemma 12 (`Omega_lower_bound_case1`). *Let $0 \le k \le E-1$ and assume $2^k \le n_E$. Then*

$$\Omega(n_E - 2^k) \ge E.$$

Proof. From $n_E \equiv 2^k \pmod{Q_k}$ we get $Q_k \mid (n_E - 2^k)$. If $n_E - 2^k = 0$ then $n_E = 2^k$, contradicting Lemma `n_E_not_pow_two`. Hence $n_E - 2^k \neq 0$ and Lemma `Omega_dvd_le` gives

$$\Omega(n_E - 2^k) \geq \Omega(Q_k) = E.$$

□

Lemma 13 (`Omega_lower_bound_case2`)**.** *Let $k \geq E$ and assume $2^k \leq n_E$. Then*

$$\Omega(n_E - 2^k) \geq E.$$

Proof. We have $2^E \mid n_E$. Also $k \geq E$ implies $2^E \mid 2^k$. Therefore $2^E \mid (n_E - 2^k)$. As before, $n_E - 2^k \neq 0$ by Lemma `n_E_not_pow_two`. Hence

$$\Omega(n_E - 2^k) \geq \Omega(2^E) = E.$$

□

Lemma 14 (`Omega_lower_bound`)**.** *For every k with $2^k \leq n_E$,*

$$\Omega(n_E - 2^k) \geq E.$$

Proof. Combine the preceding two lemmas. □

PNT-based bounds

The remaining steps control the size of n_E in terms of E (up to eventual constants), then invert that relation to obtain a lower bound on $\Omega(n_E - 2^k)$ in terms of $\sqrt{\log n_E / \log\log n_E}$.

Lemma 15 (`p_kj_bound_eventually`)**.** *There exists $C > 0$ such that for all sufficiently large E, for all $0 \leq k \leq E - 1$ and $1 \leq j \leq E$,*

$$p_{k,j} \ \leq \ C\, E^2 \log E.$$

Proof. Each $p_{k,j}$ is an n-th prime with index $n \leq E^2 + E + 1$. Apply Theorem `nth_prime_le_const_mul_log_eventually` to bound this by a constant multiple of $(E^2 + E + 1) \log(E^2 + E + 1) = O(E^2 \log E)$. □

Lemma 16 (`log_Q_k_bound_eventually`)**.** *There exists $C > 0$ such that for all sufficiently large E and all $0 \leq k \leq E - 1$,*

$$\log Q_k \ \leq \ C\, E \log E.$$

Proof. We have $\log Q_k = \sum_{j=1}^{E} \log p_{k,j}$. Using the previous lemma,

$$\log Q_k \leq E \log(C E^2 \log E) = O(E \log E).$$

□

Lemma 17 (`log_M_bound_eventually`)**.** *There exists $C > 0$ such that for all sufficiently large E,*

$$\log M \ \leq \ C\, E^2 \log E.$$

Proof. Using $M = 2^E \cdot 3 \cdot \prod_{k<E} Q_k$,

$$\log M = E \log 2 + \log 3 + \sum_{k=0}^{E-1} \log Q_k.$$

Applying the previous lemma yields the result. ☐

Lemma 18 (`n_E_lt_M`). *We have $n_E < M$.*

Proof. Follows from the CRT construction. ☐

Inversion to the $\sqrt{\log / \log\log}$ scale

Lemma 19 (`pntRate_n_E_le_const_mul_E_eventually`). *There exists $C' > 0$ such that for all sufficiently large E,*

$$\mathrm{pntRate}(n_E) \;\le\; C' E.$$

Proof. From $n_E < M$ and Lemma `log_M_bound_eventually`,

$$\log n_E \le \log M \le C E^2 \log E$$

for large E.

Since $2^E \mid n_E$ by construction, $n_E \ge 2^E$ and

$$\log n_E \ge E \log 2.$$

Thus $\log\log n_E \ge \log(E \log 2)$, which for large E is $\ge \frac{1}{2} \log E$.

Combining:

$$\mathrm{pntRate}(n_E)^2 = \frac{\log n_E}{\log\log n_E} \;\le\; \frac{C E^2 \log E}{\frac{1}{2}\log E} \;=\; 2C\,E^2,$$

giving the claim. ☐

Finishing

Theorem 1 (`main_inequality_eventually`). *There exists $c > 0$ such that for all sufficiently large E, for every k with $2^k \le n_E$,*

$$\Omega(n_E - 2^k) \;\ge\; c \cdot \mathrm{pntRate}(n_E).$$

Proof. By Lemma `Omega_lower_bound`, $\Omega(n_E - 2^k) \ge E$. By Lemma `pntRate_n_E_le_const_mul_E_eventually`, $\mathrm{pntRate}(n_E) \le C'E$. This implies the desired inequality. ☐

Definition 9 (`is_counterexample`). *Given $c > 0$, we say n is a counterexample (for that c) if for every k with $2^k \le n$,*

$$\Omega(n - 2^k) \;\ge\; c \cdot \mathrm{pntRate}(n).$$

Corollary 1 (`infinitely_many_counterexamples`). *There exists $c > 0$ such that the set of n satisfying `is_counterexample`(c, n) is infinite.*

Proof. By Theorem `main_inequality_eventually`, all sufficiently large E produce an n_E that is a counterexample for the same constant c.

Since $n_E \geq 2^E$, $n_E \to \infty$ as $E \to \infty$. Therefore the counterexamples are unbounded and infinite. $\square$