

Empirical Investigation of an Open Conjecture: Every odd integer greater than 5 can be expressed as the sum of a prime and twice

Agentic NL→Lean 4 Pipeline
Job #10

April 26, 2026

Abstract

This report documents the empirical investigation of an open mathematical conjecture that could not be formally proved or disproved in Lean 4 with Mathlib. Numerical experiments were conducted to gather evidence for or against the conjecture. The empirical verdict is: **Empirically Supported**. The conjecture remains formally open.

1 Conjecture Statement

Conjecture 1.

Every odd integer greater than 5 can be expressed as the sum of a prime and twice another prime.

2 Status

Formal Status: OPEN — no Lean 4 proof or disproof was found.

Empirical Verdict: **Empirically Supported**

The pipeline attempted formal verification in Lean 4 with Mathlib but was unable to produce a compiling proof or disproof. Empirical testing was then conducted to gather numerical evidence.

3 Basic Empirical Testing

The following output was produced by the basic numerical experiment:

```
=====
=== EXPERIMENT PLAN ===
=====

Conjecture (Lemoine / Levy, 1963):
  Every odd integer  $n > 5$  can be written as  $n = p + 2q$ ,
  where  $p$  and  $q$  are (not necessarily distinct) primes.

Strategy - test from multiple angles:
```

- (1) EXHAUSTIVE CHECK for all odd n in $(5, N_MAX]$.
Verify that at least one representation $p + 2q$ exists for each n .
If any n lacks a representation, report it as a counterexample.
- (2) COUNT the number of representations $R(n)$ for each odd $n > 5$.
Under the conjecture $R(n) \geq 1$. We will study the growth of $R(n)$
(it is conjectured to grow roughly like $n / (\ln n)^2$ on average),
which provides *redundancy evidence*: if $R(n)$ is large and
growing, random-model style heuristics strongly support the
conjecture asymptotically.
- (3) MINIMUM-PRIME analysis. For each odd n , find the smallest prime p
such that $(n-p)/2$ is prime. Track how this min- p grows with n .
A bounded, slowly growing min- p suggests the conjecture is robust;
a sudden jump would hint at possible failure.
- (4) RANDOM LARGE- n SAMPLING. Test 10,000+ large random odd integers
up to $\sim 10^7$ (not full exhaustive) to look for counterexamples
far beyond the dense regime.
- (5) EDGE / BOUNDARY CASES. Check $n = 7, 9, 11, \dots$ explicitly.
Check powers-of-two-neighbors, Mersenne-like values, large twin-
prime neighborhoods, etc.
- (6) STATISTICAL TEST. Compare empirical mean of $R(n)$ with the
heuristic prediction $n / (2 (\ln n)^2)$ (standard prime-tuples
heuristic). Correlation should be strong and positive.

If every odd $n > 5$ up to our bound has ≥ 1 representation AND $R(n)$ grows,
this is very strong empirical support (the conjecture has been verified
up to $\sim 10^{10}$ in the literature; we will re-verify a solid slice).

- [1] Exhaustive check for odd n in $(5, 200000]$...
 Odd integers tested: 99,997
 Total representation pairs (p,q) counted: 94,873,744
 Counterexamples found $(R(n)=0)$: 0
 Every odd n in $(5, 200000]$ has at least one representation.
 min $R(n) = 1$, max $R(n) = 4304$
 mean $R(n) = 948.77$, median $R(n) = 853.0$
 Max 'smallest prime p ' needed: 829 at $n=189947$
- [2] Heuristic comparison: $R(n)$ vs $n / (2 (\ln n)^2)$
 Pearson correlation (binned) between $R(n)$ and heuristic: 0.9999
 Ratio $R(n)$ /heuristic : min=2.559, max=3.125, mean=2.829
- [3] Smallest-prime-needed analysis
- | | | | |
|----------------|---------------------------|-----------------------|---------------|
| $n = 189947$: | need smallest $p = 829$, | $q = (n-p)/2 = 94559$ | (prime? True) |
| $n = 93167$: | need smallest $p = 709$, | $q = (n-p)/2 = 46229$ | (prime? True) |
| $n = 171961$: | need smallest $p = 683$, | $q = (n-p)/2 = 85639$ | (prime? True) |
| $n = 96563$: | need smallest $p = 601$, | $q = (n-p)/2 = 47981$ | (prime? True) |
| $n = 57887$: | need smallest $p = 601$, | $q = (n-p)/2 = 28643$ | (prime? True) |
| $n = 96527$: | need smallest $p = 601$, | $q = (n-p)/2 = 47963$ | (prime? True) |

```

n= 87007: need smallest p= 593, q=(n-p)/2=43207 (prime? True)
n= 54863: need smallest p= 577, q=(n-p)/2=27143 (prime? True)
n= 170099: need smallest p= 577, q=(n-p)/2=84761 (prime? True)
n= 112589: need smallest p= 571, q=(n-p)/2=56009 (prime? True)

[4] Random large-n sampling up to 107
Random large odd integers tested: 12,000 (range 7 .. 10,000,000)
Failures (no representation found): 0
Max smallest-prime p used in sample: 733
Mean smallest-prime p used in sample: 56.17

[5] Edge / boundary cases
n= 7 -> reps (first 3): [(3, 2)]
n= 9 -> reps (first 3): [(3, 3), (5, 2)]
n= 11 -> reps (first 3): [(5, 3), (7, 2)]
n= 13 -> reps (first 3): [(3, 5), (7, 3)]
n= 15 -> reps (first 3): [(5, 5), (11, 2)]
n= 17 -> reps (first 3): [(3, 7), (7, 5), (11, 3)]
n= 19 -> reps (first 3): [(5, 7), (13, 3)]
n= 21 -> reps (first 3): [(7, 7), (11, 5), (17, 2)]
n= 23 -> reps (first 3): [(13, 5), (17, 3), (19, 2)]
n= 25 -> reps (first 3): [(3, 11), (11, 7)
... [truncated]

```

4 Experiment Code (Basic)

```

import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
import numpy as np
import random
import math
from sympy import isprime, primerange, nextprime
from collections import Counter
import time

t0 = time.time()

print("=" * 70)
print("===_EXPERIMENT_PLAN_===")
print("=" * 70)
print("""
Conjecture (Lemoine / Levy, 1963):
    Every odd integer  $n > 5$  can be written as  $n = p + 2q$ ,
    where  $p$  and  $q$  are (not necessarily distinct) primes.

Strategy - test from multiple angles:

(1) EXHAUSTIVE CHECK for all odd  $n$  in  $(5, N\_MAX]$ .
    Verify that at least one representation  $p + 2q$  exists for each  $n$ .
    If any  $n$  lacks a representation, report it as a counterexample.

```

- (2) *COUNT* the number of representations $R(n)$ for each odd $n > 5$. Under the conjecture $R(n) \geq 1$. We will study the growth of $R(n)$ (it is conjectured to grow roughly like $n / (\ln n)^2$ on average), which provides **redundancy evidence**: if $R(n)$ is large and growing, random-model style heuristics strongly support the conjecture asymptotically.
- (3) *MINIMUM-PRIME* analysis. For each odd n , find the smallest prime p such that $(n-p)/2$ is prime. Track how this $\text{min-}p$ grows with n . A bounded, slowly growing $\text{min-}p$ suggests the conjecture is robust; a sudden jump would hint at possible failure.
- (4) *RANDOM LARGE- n SAMPLING*. Test 10,000+ large random odd integers up to $\sim 10^7$ (not full exhaustive) to look for counterexamples far beyond the dense regime.
- (5) *EDGE / BOUNDARY CASES*. Check $n = 7, 9, 11, \dots$ explicitly. Check powers-of-two-neighbors, Mersenne-like values, large twin-prime neighborhoods, etc.
- (6) *STATISTICAL TEST*. Compare empirical mean of $R(n)$ with the heuristic prediction $n / (2 (\ln n)^2)$ (standard prime-tuples heuristic). Correlation should be strong and positive.

If every odd $n > 5$ up to our bound has ≥ 1 representation AND $R(n)$ grows, this is very strong empirical support (the conjecture has been verified up to $\sim 10^{10}$ in the literature; we will re-verify a solid slice).

""")

```
# ----- (1) EXHAUSTIVE CHECK -----
N_MAX_EXH = 200_000
print(f"\n[1] Exhaustive check for odd n in (5, {N_MAX_EXH}]...")

# Sieve of Eratosthenes up to N_MAX_EXH
sieve = np.ones(N_MAX_EXH + 1, dtype=bool)
sieve[:2] = False
for i in range(2, int(math.isqrt(N_MAX_EXH)) + 1):
    if sieve[i]:
        sieve[i*i::i] = False
is_prime = sieve # boolean array

# Primes <= N_MAX_EXH
primes = np.nonzero(is_prime)[0]

# For each odd n > 5, count representations n = p + 2q
# We iterate q over primes with 2q < n, then check if n - 2q is prime.
# Vectorize per-n via accumulation:
odd_ns = np.arange(7, N_MAX_EXH + 1, 2)
R = np.zeros_like(odd_ns, dtype=np.int64)
min_p = np.full_like(odd_ns, -1, dtype=np.int64) # smallest p such that (n-
p)/2 is prime

# For each prime q, mark all odd n = p + 2q where p is prime.
```

```

# Equivalent: for each prime q, for each prime p (odd p or p=2), n = p + 2q.
# Iterate over q; then n - 2q must be prime (=p).
# We'll do it per-n by checking primes q up to (n-2)/2.

# Faster: for each n, iterate primes p in increasing order, test (n-p)/2 is
  prime.
# Use a boolean lookup `is_prime` for (n-p)/2 which is <= (n-2)/2 <=
  N_MAX_EXH/2.

# Precompute small-prime list (all primes up to N_MAX_EXH-4):
primes_list = primes.tolist()

# To keep it fast & vectorized: for each prime p, mark all odd n >= p+4 such
  that
# (n-p)/2 is prime. That is, n = p + 2q for every prime q >= 2.
# We'll loop over primes p (a lot of them), and for each p loop over q? Too
  slow.
# Instead: loop over p (outer) and q (inner over primes list), set R[n]+=1
  and min_p.
# Number of pairs ~ pi(N)^2 / 2 which is ~9600^2 = 9.2e7. Borderline.
# We'll use numpy: for each prime p, compute all n = p + 2*primes_array <=
  N_MAX_EXH.

primes_arr = primes.astype(np.int64)

# map from odd n to index in odd_ns: idx = (n-7)//2
for p in primes_arr:
    ns = p + 2 * primes_arr
    ns = ns[(ns <= N_MAX_EXH) & (ns >= 7) & (ns % 2 == 1)]
    idx = (ns - 7) // 2
    # Count each representation
    np.add.at(R, idx, 1)
    # Track minimum p
    # For entries where min_p == -1 or p < min_p, set min_p = p
    current = min_p[idx]
    mask = (current == -1) | (p < current)
    min_p[idx[mask]] = p

missing = odd_ns[R == 0]
print(f"Odd integers tested: {len(odd_ns):,}")
print(f"Total representation pairs (p,q) counted: {R.sum():,}")
print(f"Counterexamples found (R(n)==0): {len(missing)}")
if len(missing) > 0:
    print(f"First counterexamples: {missing[:20].tolist()}")
else:
    print(f"Every odd n in (5, {N_MAX_EXH}] has at least one
      representation.")

print(f"min R(n) = {R.min()}, max R(n) = {R.max()}")
print(f"mean R(n) = {R.mean():.2f}, median R(n) = {np.median(R):.1f}")
print(f"Max 'smallest prime p' needed: {min_p.max()} at n={odd_ns[np.
  argmax(min_p)]}")

# ----- (2) R(n) vs heuristic -----

```

```
print("\n[2] Heuristic comparison:  $R(n)$  vs  $n / (2(\ln n)^2)$ ")  
# ... [truncated]
```

5 Conclusion

The conjecture remains formally open. Numerical experiments **support** the conjecture — no counterexamples were found across all tested parameter ranges. Further investigation (both formal and empirical) is warranted.