

5 machine-checked refutations of machine-generated conjectures

Augusto Rehfeldt

Text generated by language models in the mathforge pipeline; not reviewed by a person.

Abstract

We record 5 false statements, each with an explicit counterexample. The statements were conjectured by a language model inside an automated pipeline (mathforge), not taken from the literature. For each, a search script found a witness, an independently written script re-checked it from the statement alone, and Lean 4 with Mathlib accepted a sorry-free proof of the negation, whose statement a separate model call judged faithful to the informal one. No person has reviewed the statements, the witnesses or the Lean files.

1. Introduction

A counterexample is a small true fact about the objects, and it is normally thrown away. The claims below are plausible-looking statements over infinite families that fail at a small instance. They are collected so that nobody has to rediscover that they fail. Each section gives the claim, the witness, how it was checked, and a link to the scripts and the Lean file.

2. Refutations

2.1 Opposite-end strip animals occur in even numbers unless their area and perimeter match two parity conditions

Seed topic of the run: Connected cell sets in a $2 \times n$ strip that meet both end columns, refined by area, perimeter, and the occupancy of each end column; look for column-extension recurrences and reflection-based parity congruences.

Notation. Cells sharing an edge are adjacent; connected means connected under this adjacency. Area is the number of cells, and perimeter is the number of cell edges bordering an unoccupied cell or the exterior. A column contains only its top or bottom cell when exactly that cell is occupied.

Proposition 1. *The following statement is false. For every integer $n \geq 2$ and all integers $a, p \geq 0$, let $C(n, a, p)$ count connected cell sets in a $2 \times n$ strip whose first column contains only its top cell, whose last column contains only its bottom cell, and whose area and perimeter are a and p . Then $C(n, a, p)$ is even whenever a is odd or p is not congruent to $2(n-1)$ modulo 4.*

Proof. Counterexample: $n=3$, $a=4$, $p=10$; $C(n, a, p)=1$, actual parity=1, claimed parity=0. An independent script, written by a separate model call that re-implemented the definitions from the statement, confirmed the witness.

□

Record of the checks:

- Lean 4 + Mathlib: compiled, sorry-free (0 occurrences of the token in the source, none reported by the elaborator).

- Back-translation of the Lean statement: **FAITHFUL**. The theorem says it is false that, for every integer $n \geq 2$ and nonnegative integers a and p , the number of connected occupied subsets of a $2 \times n$ strip with a top-only first column, a bottom-only last column, area a , and perimeter p is even whenever a is odd or p is not congruent to $2(n-1)$ modulo 4.
- Scripts, Lean file and Lake project: <https://github.com/augusto-rehfeldt/mathforge-results/tree/main/n/connected-cell-sets-in-a-2-n-strip-that-meet-bot-c1>
- Models: gpt-6-sol

2.2 Compressing two prime-sized residue classes preserves partition counts modulo prime squares, apart from an explicit paired-block correction

Seed topic of the run: Set partitions of $[n]$ in which every block has element sum divisible by three, refined by the numbers of blocks and singleton blocks; look for residue-class deletion recurrences and coefficient congruences induced by cyclically permuting labels within each residue class.

Notation. $[n] = \{1, \dots, n\}$. For any finite set S of positive integers, $F_S(x, y)$ is the sum of $x^k y^s$ over all set partitions of S whose every block has element sum divisible by 3, where k is the number of blocks and s is the number of singleton blocks. The empty partition contributes 1. Introduce two distinct formal objects a and b , not belonging to R , with weights $w(a)=p$ and $w(b)=2p$; set $w(r)=r$ for $r \in R$. $H_{R,p}(x, y)$ is the sum of $x^k y^t$ over all partitions of $R \cup \{a, b\}$ whose every block has total weight divisible by 3, where k counts all blocks and t counts only singleton blocks $\{r\}$ with $r \in R$. Thus singleton blocks containing a or b receive no factor y . The symbols x and y are commuting indeterminates, and $p! = 1 \cdot 2 \cdot \dots \cdot p$. Coefficientwise congruence means that every coefficient of the difference is divisible by p^2 .

Proposition 2. *The following statement is false. For every prime $p \geq 5$ and every integer $n \geq 3p-1$, let A consist of the p smallest elements of $[n]$ congruent to 1 modulo 3, let B consist of the p smallest elements congruent to 2 modulo 3, and put $R = [n] \setminus (A \cup B)$. Then, coefficientwise in the integer polynomial ring $\mathbb{Z}[x, y]$, $F_{[n]}(x, y) \equiv H_{R,p}(x, y) + p! x^p F_R(x, y)$ modulo p^2 .*

Proof. Counterexample: $p=5$, $n=14$, coefficient $x^2 y^0$, left=2731, right=11, residues modulo 25: 6, 11. An independent script, written by a separate model call that re-implemented the definitions from the statement, confirmed the witness.

□

Record of the checks:

- Lean 4 + Mathlib: compiled, sorry-free (0 occurrences of the token in the source, none reported by the elaborator).
- Back-translation of the Lean statement: **FAITHFUL**. The theorem denies the following universal assertion: for all natural numbers p and n , if p is prime, $p \geq 5$, and $n \geq 3p-1$, then for every pair of natural numbers k, t , p^2 divides the difference between two integer coefficients. The first counts partitions of $\{1, \dots, n\}$ into k nonempty blocks with each block's sum divisible by 3 and exactly t singleton blocks. The second counts such weighted partitions of the remaining elements together with two distinct added items of weights p and $2p$, counting only ordinary singleton blocks toward t , and adds $p!$ times the remaining-set partition count with $k-p$ blocks and t singletons when $k \geq p$. The remaining elements exclude residue-1 elements at most $3p-2$ and residue-2 elements at most $3p-1$.
- Scripts, Lean file and Lake project: <https://github.com/augusto-rehfeldt/mathforge-results/tree/main/n/set-partitions-of-n-in-which-every-block-has-ele-c4>
- Models: gpt-6.1-sol

2.3 Odd-length partition counts modulo two reduce to weighted matchings on blocks of half-length partitions

Seed topic of the run: Set partitions of $[n]$ in which no block contains two elements differing by two, refined by the number of blocks, singleton blocks, and indices i whose consecutive elements i and $i+1$ lie in the same

block; look for insertion recurrences tracking the last two block memberships and reversal-induced coefficient congruences.

Notation. $[r]=\{1,\dots,r\}$. A partition is admissible if no block contains two integers differing by two. For an admissible partition P , $b(P)$ is its number of blocks, $s(P)$ its number of singleton blocks, and $a(P)=|\{i\in[r-1]: i \text{ and } i+1 \text{ belong to the same block of } P\}|$. Define $F_r(x,y,z)=\sum_P x^{b(P)} y^{s(P)} z^{a(P)}$, summing over admissible partitions of $[r]$; x,y,z are commuting indeterminates. In the displayed congruence, $b=b(Q)$. A matching on a finite set is a collection of pairwise disjoint unordered pairs of distinct members, including the empty collection. An absent D removes no block. Coefficientwise congruence means that every corresponding pair of integer coefficients has even difference.

Proposition 3. *The following statement is false. For every integer $m \geq 2$, the following polynomial congruence holds coefficientwise modulo 2: $F_{2m+1}(x,y,z) \equiv \sum_Q \sum_{\{D,I,M\}} x^{2b-|I|-2\delta+1} y^{2u+1-\delta} z^{2a(Q)}$. The outer sum ranges over all admissible partitions Q of $[m]$. In each inner sum, D is either absent or a block of Q containing neither $m-1$ nor m ; δ is 0 or 1 according as D is absent or present. The set I is any subset of the blocks other than D that excludes the block containing m . The object M is any matching on the blocks outside I and D . The integer u counts singleton blocks of Q outside I and D that are unmatched in M .*

Proof. Counterexample: $m=2$, $\text{exponents}=(3, 1, 0)$; $\text{LHS}=1 \pmod{2=1}$, $\text{RHS}=0 \pmod{2=0}$. An independent script, written by a separate model call that re-implemented the definitions from the statement, confirmed the witness. □

Record of the checks:

- Lean 4 + Mathlib: compiled, sorry-free (0 occurrences of the token in the source, none reported by the elaborator).
- Back-translation of the Lean statement: **FAITHFUL**. The theorem denies that, for every natural number $m \geq 2$ and every triple of natural exponents p,q,t , the number of admissible partitions of $[2m+1]$ with p blocks, q singleton blocks, and t adjacent pairs in a common block has the same parity as the following count: admissible partitions Q of $[m]$, choices of an absent distinguished block or a block D containing neither $m-1$ nor m , subsets I of the other blocks excluding the block containing m , and matchings M on the remaining blocks, whose exponents are $p = 2|Q|-|I|-2\delta+1$, $q = 2u+1-\delta$, and $t = 2a(Q)$, where u counts unmatched singleton blocks. The statement inside the negation asserts this parity equality universally.
- Scripts, Lean file and Lake project: <https://github.com/augusto-rehfeldt/mathforge-results/tree/main/set-partitions-of-n-in-which-no-block-contains-t-c2>
- Models: gpt-6.1-sol

2.4 Every self-complementary Sidon family with given sum and maximum has even cardinality

Seed topic of the run: Sidon (B_2) subsets of $\{1,\dots,n\}$ and of cyclic groups Z_n , with all pairwise sums distinct, refined by cardinality, element sum, and maximal element, looking for enumerative identities, recurrences across n , modular counting congruences, and translation, reflection, and dilation equinumerosity relations together with interval-to-cyclic lifting identities.

Notation. n is the interval bound; $\{1,\dots,n\}$ is the ground set; k is cardinality; s = sum of elements of A ; m = largest element of A ; Sidon means the $k^*(k+1)/2$ sums $a+b$ with $a \leq b$ are all distinct; $N(n,k,s,m)$ counts such sets A .

Proposition 4. *The following statement is false. For all integers $n \geq 1$, $k \geq 3$, $m \geq 1$, and $s \geq 1$ with $m \leq n$, if $2^*s = k^*(m+1)$ then $N(n,k,s,m)$ is even, where $N(n,k,s,m)$ is the number of subsets A of $\{1,\dots,n\}$ with $|A| = k$, $\text{sum}(A) = s$, $\text{max}(A) = m$, and all sums $a+b$ with a,b in A and $a \leq b$ pairwise distinct.*

Proof. Counterexample: `n=7 k=3 s=12 m=7 2*s=24 k*(m+1)=24 N=1 N_even=False sets=[(2, 3, 7)]`. An independent script, written by a separate model call that re-implemented the definitions from the statement, confirmed the witness.

□

Record of the checks:

- Lean 4 + Mathlib: compiled, sorry-free (0 occurrences of the token in the source, none reported by the elaborator).
- Back-translation of the Lean statement: **FAITHFUL**. For all natural numbers n, k, s, m with $n \geq 1, k \geq 3, m \geq 1, s \geq 1$ and $m \leq n$, if $2*s = k*(m+1)$ then $N(n,k,s,m)$ is even, where $N(n,k,s,m)$ is the number of k -element subsets A of $\{1, \dots, n\}$ such that the sum of A equals s , the supremum of A equals m , and A satisfies IsSidon, meaning the sums $a+b$ for pairs (a,b) in $A \times A$ with $a \leq b$ are all pairwise distinct.
- Scripts, Lean file and Lake project: <https://github.com/augusto-rehfeldt/mathforge-results/tree/main/sidon-b-2-subsets-of-1-n-and-of-cyclic-groups-z--c4>
- Models: muse-spark-1.3-contributor-free

2.5 For matrices of order at least six, only two inverse weights sit closest below the maximum

Seed topic of the run: Upper unitriangular matrices over F_2 whose first superdiagonal is all ones and whose second superdiagonal is arbitrary, refined by the number of nonzero entries in their inverses; look for end-extension recurrences and reversal-symmetry congruences.

Notation. F_2 is the field with two elements. The second superdiagonal consists of entries $A_{\{i,i+2\}}$ for $1 \leq i \leq n-2$; its first and last positions have $i=1$ and $i=n-2$, respectively. Inverse weight counts nonzero entries $(A^{-1})_{\{i,j\}}$ with $i < j$.

Proposition 5. *The following statement is false. For every integer $n \geq 6$, let A range over the $n \times n$ upper unitriangular matrices over F_2 whose first superdiagonal is all ones and whose entries outside the diagonal and first two superdiagonals are zero. If A is not the matrix with zero second superdiagonal, then its inverse has at most $n(n-1)/2 - (n-2)$ nonzero entries above the diagonal. Equality holds exactly when the second superdiagonal has a single one, at its first or last position.*

Proof. Counterexample: `n=6, second superdiagonal=[1, 1, 0, 0]; inverse weight=11, claimed bound=11`. An independent script, written by a separate model call that re-implemented the definitions from the statement, confirmed the witness.

□

Record of the checks:

- Lean 4 + Mathlib: compiled, sorry-free (0 occurrences of the token in the source, none reported by the elaborator).
- Back-translation of the Lean statement: **FAITHFUL**. It is false that, for every natural number $n \geq 6$, every admissible $n \times n$ matrix over $\mathbb{Z} \bmod 2$ with a nonzero entry on its second superdiagonal, and every right inverse B of that matrix, the number of nonzero entries of B above the diagonal is at most $n(n-1)/2 - (n-2)$, with equality exactly when the matrix has a single nonzero second-superdiagonal entry at its first or last position.
- Scripts, Lean file and Lake project: <https://github.com/augusto-rehfeldt/mathforge-results/tree/main/upper-unitriangular-matrices-over-f-whose-first--c2>
- Models: gpt-6-sol

3. Limitations

Every statement, script and Lean file here was written by language models in an automated pipeline; no person has reviewed them. Lean certifies the Lean statement only. Whether that statement says what the informal one says was judged by a separate model call that back-translated it, which is a screening filter, not an oracle, so each Lean file should be read against its statement.

The conjectures were posed by the pipeline itself, so their refutation carries no claim of significance. No literature search was run for these statements: any of them, or its failure, may already be known.