Compact trie nodes and pointers for random marked involutions

TAOCP exercise 7.2.2.4–205

Shisheng Li*

May 2026

Abstract

For the compact trie of TAOCP exercise 7.2.2.4–204 built from n i.i.d. random length-q marked involutions, we prove, as $q \to \infty$ with $n \geq 1$ fixed,

$$P_{\text{nodes}}(n,q) = nq + (n-1) - c_n^* \sqrt{q} - d_n^* + \mathrm{O}(q^{-1/2}),$$
$$P_{\text{ptr}}(n,q) = \tfrac{n}{6}q^2 - \tfrac{n}{3}q^{3/2} + A_n q + B_n\sqrt{q} + \mathrm{O}(1),$$

where $c_j := \int_0^\infty e^{-(j/2)\xi^2-2j\xi}\,d\xi = \sqrt{\pi/(2j)}\,e^{2j}\,\mathrm{erfc}(\sqrt{2j})$, the constants c_n^*, d_n^*, B_n are explicit (rational + $\mathbb{Q}$-linear in $c_1,\ldots,c_n$) combinations (c_n^*, d_n^* involve only $c_2,\ldots,c_n$; c_1 enters B_n via the term $n\,c_1$), and $\boxed{A_n = \frac{5n}{2} + H_n + 2c_n^*}$, a clean closed form mediated by the classical alternating-binomial identity $\sum_{j=1}^n (-1)^{j+1}\binom{n}{j}/j = H_n$. For general n we leave the O(1) residual of $P_{\text{ptr}}(n,q)$ implicit (its rational + $\mathbb{Q}$-linear-in-c_j structure is captured by Cor. 21); at $n = 1$ we resolve it to the fifth term: $P_{\text{ptr}}(1,q) = q^2/6 - q^{3/2}/3 + 7q/2 + (c_1 - 31/12)\sqrt{q} + (17/12 + c_1) + \mathrm{O}(q^{-1/2})$. All closed forms are verified against exact-rational enumeration up through $(n,q) = (4,5)$ for P_{nodes} and $(n,q) = (4,4)$ for P_{ptr}, and against log-space float64 evaluation through $q = 10^4$. The only non-elementary analytic input is the saddle-point asymptotic of T_N (the marked-involution count, OEIS A005425), in the form of Knuth's exercise 7.2.2.4–177.

1 Introduction

In the typescript of Pre-Fascicle 8A (8 May 2026 revision), §7.2.2.4 "Hamiltonian paths and cycles: dynamic enumeration", Knuth introduces *Algorithm E* (DYNAHAM), a frontier-based dynamic-programming algorithm that, given a graph G on $\{1,\ldots,n\}$, computes the cycle counts CYC$[m]$ of every induced subgraph $G_m = G|\{1,\ldots,m\}$ for $3 \leq m \leq n$. As Algorithm E advances from level $m-1$ to level m, its state is an *m-class* — an equivalence class of edge sets satisfying three combinatorial constraints — encoded, by the convention of exercise 178, as a string $a_1 \ldots a_q$ of length $q = |F_m|$ over the alphabet $\{\bar{1}, 0, 1, 2, \ldots, \lfloor q/2 \rfloor\}$. Such a string is precisely a marked involution of order q in the sense of that exercise. The set of m-classes that arise during the algorithm is stored in a trie keyed by these strings; exercise 204 designs a *compact* version of this trie, in which each node stores only those pointer fields that correspond to feasible extensions consistent with the prefix's open-pair structure and the remaining length. The construction allocates, at a node of length l holding TMS ("two-marked-stack") open pairs — i.e. k_p in the notation of §3; `tms` is the matching CWEB variable name in DYNAHAM —

$$d(l,\mathrm{TMS}) = \begin{cases} \mathrm{TMS} & \text{if } l + \mathrm{TMS} = q, \\ \mathrm{TMS}+2 & \text{if } l + \mathrm{TMS} = q-1, \\ \mathrm{TMS}+3 & \text{otherwise,} \end{cases} \tag{1}$$

*daizisheng@gmail.com

field positions, each holding a pointer to a child or to null (see Knuth's answer to exercise 204).[1] Throughout the remainder of the paper we write k_p for the open-pair count (denoted TMS in (1) and `tms` in the CWEB), and abbreviate $d_p := d(l_p, k_p)$ with $l_p = |p|$. The total *pointer count* of the trie is the sum of d_p over all distinct prefix nodes.

Exercise 205, rated [HM46], asks:

> *Analyze the average total number of pointers, when the method of exercise 204 builds a compact trie from n random keys $a_1 \dots a_q$ (that is, from n independently generated marked involutions as in exercise 176, with repetitions allowed).*

The printed answer is a single sentence:

> *The simpler problem for n fixed at 1 should presumably be solved first, with its asymptotics when $q \to \infty$. The average number of trie nodes (distinct prefixes of n random keys) is also of interest.*

The printed answer references two distinct quantities: the "average total number of pointers" $P_{\text{ptr}}(n,q)$, and the "average number of trie nodes" $P_{\text{nodes}}(n,q)$. With d_p defined by (1),

$$P_{\text{ptr}}(n,q) \;=\; \mathbf{E}\Big[\sum_{p\in\text{Trie}} d_p\Big], \qquad P_{\text{nodes}}(n,q) \;=\; \mathbf{E}\Big[\sum_{\substack{p\in\text{Trie}\\ |p|\ge 1}} 1\Big],$$

where the expectation is over n i.i.d. uniform marked involutions of order q, and Trie is the set of nodes of the compact trie of those keys. We do not count the root in P_{nodes}, matching the $\sum_{l=1}^{q}$ convention used throughout.

This paper gives closed-form asymptotic expansions of both quantities for every fixed $n \ge 1$: a four-term expansion of $P_{\text{nodes}}(n,q) = nq + (n-1) - c_n^*\sqrt{q} - d_n^* + \mathrm{O}(q^{-1/2})$ (Theorem 13, with c_n^*, d_n^* explicit inclusion–exclusion combinations of the boundary-layer integrals $c_j := \int_0^\infty e^{-(j/2)\xi^2 - 2j\xi}\, d\xi$) and a four-term expansion of $P_{\text{ptr}}(n,q)$ through order $\sqrt{q}$ (Theorem 19). The trivial-looking case $n = 1$ already contributes $P_{\text{nodes}}(1,q) = q$ (one key forms a single path with q distinct prefixes) but $P_{\text{ptr}}(1,q) \sim q^2/6$ — on average, each prefix node carries $\Theta(\sqrt{q})$ field positions, so the total pointer count grows quadratically in q even for one key. The interesting structure for P_{nodes} appears at $n \ge 2$, where prefix sharing between keys reduces the count below nq by a $-c_n^*\sqrt{q}$ amount.

The generating-function pair attached to marked involutions [1, exercise 7.2.2.4–177] is

$$\sum_{q\ge 0} T_q \,\frac{z^q}{q!} \;=\; \exp\Big(2z + \tfrac{z^2}{2}\Big), \tag{2}$$

where $T_q := |\{\text{length-}q \text{ marked involutions}\}|$ equals $a(q)$ of OEIS A005425 and satisfies the recurrence $T_{N+2} = 2T_{N+1} + (N+1)T_N$ with initial values $T_0 = 1$, $T_1 = 2$:

[1] Equation (1) encodes exactly what the CWEB section `<Allocate a new node for level l+1>` in Dynaham [2] computes. The CWEB variable `l` holds the *parent* level, so the new node being analysed sits at level $l = \texttt{l} + 1$ in the paper's notation; with TMS = `tms` its open-pair count, the three branches of the code map one-to-one onto the three cases of (1):

- `else` branch ($\texttt{l} + 1 + \texttt{tms} \ge q$, i.e. $l + \text{TMS} = q$): `slots = tms`, total TMS fields (case 1).
- `if`-branch with `l+2+tms==q` (i.e. $l + \text{TMS} = q - 1$): two extra fields plus `slots` = `tms`, total TMS + 2 fields (case 2).
- `if`-branch with `l+2+tms<q` (i.e. $l + \text{TMS} \le q - 2$): two extra fields plus `slots` = `tms` + 1, total TMS + 3 fields (case 3, "otherwise").

The two extra fields hold the codes 0 and $\bar{1}$, present whenever there is room beneath the node; the `slots` positive entries $1, \dots,$ TMS are augmented by one extra slot in case 3 to accommodate opening a new pair.

N	0	1	2	3	4	5	6	7	8	9
T_N	1	2	5	14	43	142	499	1850	7193	29186

(verified against OEIS A005425's b-file). The factor $2z$ is the two-flavour 1-cycle (marked or unmarked), the $z^2/2$ is the 2-cycle. Compare with the standard (unmarked) involution count I_N (OEIS A000085; $1, 1, 2, 4, 10, 26, \dots$), which has EGF $\exp(z + z^2/2)$ and recurrence $I_{N+2} = I_{N+1} + (N+1)I_N$; the marked variant differs only in the linear factor $2z$ vs. z of the EGF, but this changes the Wright asymptotic constant and the boundary-layer profile $\exp(-\xi^2/2-2\xi)$ specific to exercise 205.

Throughout the paper, n denotes the (fixed) number of random keys of exercise 205, while the running index of the T-sequence and of Wright's asymptotic is written N to avoid collision.

Background reading. The classical analysis of trie height and pointer count under the uniform i.i.d. model is in Knuth's TAOCP Vol. 3 [5] and Szpankowski [7]; the analytic-combinatorics framework, including Mellin-transform techniques for digital-tree asymptotics, is developed in Flajolet–Régnier–Sedgewick [6]; the involution-counting recurrence and EGF (with weight 1 per fixed point) appears in Riordan [8].

2 Main results and structure of the paper

Main results. For every fixed integer $n \geq 1$, as $q \to \infty$,

$$\boxed{\begin{aligned} P_{\text{nodes}}(n,q) &= n\,q + (n-1) - c_n^*\sqrt{q} - d_n^* + \mathrm{O}(q^{-1/2}), \\ P_{\text{ptr}}(n,q) &= \tfrac{n}{6}\,q^2 - \tfrac{n}{3}\,q^{3/2} + A_n\,q + B_n\sqrt{q} + \mathrm{O}(1), \end{aligned}}$$

with closed-form coefficients

$$c_n^* = \sum_{j=2}^{n}(-1)^j\binom{n}{j}c_j, \qquad c_j := \int_0^\infty e^{-(j/2)\xi^2-2j\xi}\,d\xi,$$

$$A_n = \tfrac{5n}{2} + H_n + 2c_n^* \quad \text{(Theorem 20)}, \qquad B_n = n\big(c_1 - \tfrac{31}{12}\big) - 3c_n^* + \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}d_j^{(k)}.$$

Glossary. $c_n^* = \sum_{j=2}^{n}(-1)^j\binom{n}{j}c_j$ (boundary-layer integral combination, appearing in both P_{nodes} and A_n, B_n); $d_n^* = \sum_{j=2}^{n}(-1)^j\binom{n}{j}\mathrm{const}_j$ (parallel combination of sub-leading constants); $c_j^{(k)}, d_j^{(k)}$ are the leading and sub-leading coefficients of $S_j^{(k)}(q)$ respectively (Lemma 18). All four families are $\mathbb{Q}$-linear in $\{c_j\}$ with rational shifts. The recurrence of c_n^* across P_{nodes}'s $\sqrt{q}$-term and P_{ptr}'s q- and $\sqrt{q}$-terms — combined with the classical identity $\sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}/j = H_n$ — is what produces the clean $A_n = 5n/2 + H_n + 2c_n^*$ of Theorem 20. The $n = 1$ specialisation is $P_{\text{nodes}}(1,q) = q$ and $P_{\text{ptr}}(1,q) = \frac{q^2}{6} - \frac{q^{3/2}}{3} + \frac{7q}{2} + (c_1 - \frac{31}{12})\sqrt{q} + (\frac{17}{12} + c_1) + \mathrm{O}(q^{-1/2})$.

Three building blocks. Both expansions reduce to asymptotics of three families of sums over a marked involution of size q:

- the *prefix moments* $S_j(q) = 1 + \sum_{l\geq 1,k} f(l,k)[h(q-l,k)/T_q]^j$ (the "+1" is the root $l = 0$ contribution), contributing to P_{nodes} at order $\sqrt{q}$ (Lemma 12: $S_j = c_j\sqrt{q} + \mathrm{const}_j + \mathrm{O}(q^{-1/2})$);
- the *k-weighted prefix moments* $S_j^{(k)}(q) = \sum_{l,k} kf(l,k)[h(q-l,k)/T_q]^j$, contributing to P_{ptr} at orders $q^2, q^{3/2}, q, \sqrt{q}$ (Lemma 14 for $j = 1$, Lemma 18 for $j \geq 2$);
- the *boundary moments* $B_j^{(b)}(q)$ collected on the line $q - l - k = b$, contributing to P_{ptr} at $\sqrt{q}$ (Lemma 15 for $j = 1$, Lemma 17 for $j \geq 2$ — for fixed $j \geq 2$ these decay like $\exp(-(j\log 2)/2 \cdot q\,(1 + \mathrm{o}(1)))$, the $j = 2$ term $(\exp(-q\log 2\,(1 + \mathrm{o}(1))))$ dominating the collapse of $R(n,q)$ to $n\,R(1,q)$).

The unifying analytical input is the *boundary-layer profile* $h(q-k,k)/T_q = e^{-\xi^2/2-2\xi}(1+O((\xi^3+1)/\sqrt{q}))$ for $\xi = k/\sqrt{q}$ on $k \le \sqrt{q \log q}$ (Lemma 9), itself derived from the Wright/Hayman saddle-point asymptotic of T_N.

Structure of the paper. The paper splits into two parts.

Part I (§3–§5): combinatorial reductions. §3 defines $f(l,k)$, $h(r,k)$, and proves their closed forms (Lemmas 2, 3); §3 also derives the keystone identity (Lemma 4) and the closed form for $S_1^{(k)}$ (Lemma 5). §4 expresses both P_{nodes} and P_{ptr} as finite linear combinations of S_j, $S_j^{(k)}$, and $B_j^{(b)}$ by inclusion–exclusion (Theorems 6, 7), and gives a closed form for $R(1,q)$ (Proposition 8). Part I is exact: no asymptotics enter until Part II. §5 recapitulates.

Part II (§6–§10): asymptotic analysis. §6 establishes the boundary-layer profile (Lemma 9) and the leading S_j asymptotic (Lemma 12). §7 assembles P_{nodes} (Theorem 13). §8 handles the $n = 1$ pointer count via the asymptotics of $W(1,q) = S_1^{(k)}(q)$ (Lemma 14) and $R(1,q)$ (Lemma 15), yielding the four-term expansion of $P_{\text{ptr}}(1,q)$ (Theorem 16). §9 extends to general n: the boundary moments collapse (Lemma 17), the k-weighted moments admit a two-term closed form (Lemma 18), and assembling everything proves Theorem 19. §10 recapitulates the analytical content.

Verification. Every claim is checked numerically against either exact-rational enumeration (small q) or log-space float64 evaluation of the defining sums (up to $q = 10^4$). Verification tables appear inline after every Lemma, Theorem, and Proposition.

Part I
Combinatorial reductions

The first part of this paper develops the algebraic and combinatorial identities that reduce $P_{\text{nodes}}(n,q)$ and $P_{\text{ptr}}(n,q)$ to per-prefix moments and boundary sums. Everything in Part I holds as *exact* identities for every $n \ge 1$ and $q \ge 0$; no asymptotics are used until Part II.

Verification methodology. Every closed form in Part I is accompanied by exact-rational verification against direct enumeration, implemented in Rust with the `num-bigint` and `num-rational` crates. The common ground-truth object is the explicit set $\mathcal{S}_q$ of all length-q marked involutions, built by enumerating pair partitions of $\{1,\ldots,q\}$ and marking each fixed point as $\bar{1}$ or 0 (*without* using the recurrence $T_{N+2} = 2T_{N+1} + (N+1)T_N$), then computing the canonical string of exercise 178 for each. Each verification below states its own specific enumeration method on $\mathcal{S}_q$; the column "closed form" evaluates the formula in `BigInt`/`BigRational`; ✓ marks digit-for-digit equality.

3 Notation and preliminaries

A *marked involution* of order q is a permutation σ of $\{1,\ldots,q\}$ with $\sigma^2 = \text{id}$, in which each fixed point is marked or unmarked. Following Knuth's exercise 178, we encode it as a string $a_1 \ldots a_q$ with each $a_j \in \{\bar{1}, 0, 1, 2, \ldots, \lfloor q/2 \rfloor\}$, where $\bar{1}$ marks a marked 1-cycle, 0 an unmarked 1-cycle, and a positive label t marks the two endpoints of a single 2-cycle (the smaller endpoint receives a fresh label from a strictly increasing reservoir).

Definition 1 (Prefix counts)**.** For $0 \le k \le l$, let $f(l,k)$ be the number of length-l prefixes of marked involutions in which exactly k pair labels remain *open* (their partner has not yet appeared). For $0 \le k \le r$, let $h(r,k)$ be the number of completions of a k-open prefix into a full marked involution after r further positions.

Lemma 2 (Closed form for f)**.** $f(l,k) = \binom{l}{k} T_{l-k}$.

Proof. Figure 1 illustrates the bijection.

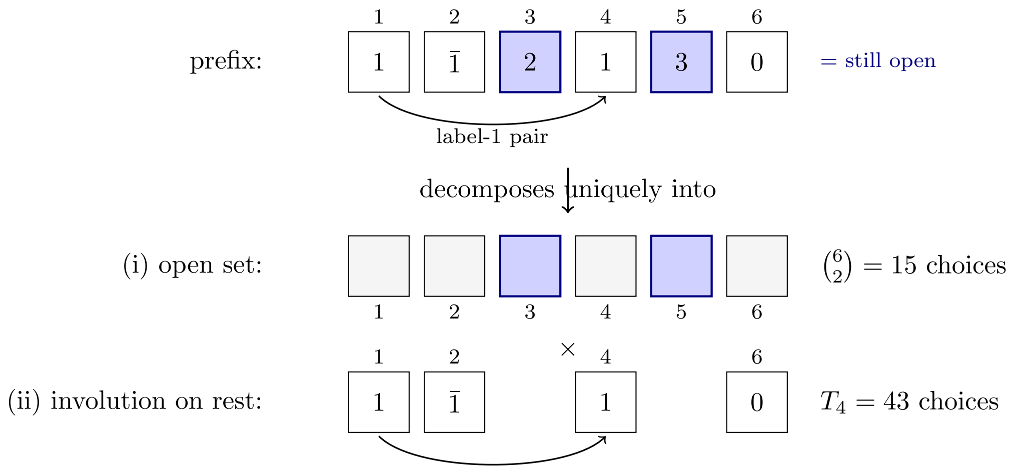

Figure 1: The bijection for Lemma 2: a length-l prefix with k still-open pairs decomposes uniquely into the *set* of k open-left-endpoint positions and a marked involution on the remaining $l-k$ positions. Once these two data are fixed, the canonical encoding rule of exercise 178 forces every label, so no further choice remains. Example shown: $l = 6$, $k = 2$, with open positions $\{3, 5\}$ and the marked involution $1 \leftrightarrow 4$, $2 = \bar{1}$, $6 = 0$ on $\{1, 2, 4, 6\}$. Total: $\binom{6}{2} \cdot T_4 = 15 \cdot 43 = 645 = f(6, 2)$.

A length-l prefix with exactly k open pairs at position l is uniquely determined by (i) the set of k positions among $\{1, \ldots, l\}$ that hold the *left* endpoints of the still-open pairs ($\binom{l}{k}$ choices); and (ii) a marked involution on the remaining $l-k$ positions, accounting for all closed 2-cycles together with marked and unmarked 1-cycles (T_{l-k} choices). Once these two data are fixed, the labels at the k open positions and at the closed-pair endpoints are forced by the canonical encoding rule of exercise 178 (each new opening takes the smallest positive integer not yet appearing in the prefix so far; closed-pair endpoints repeat their opening label), so no further choice remains. Hence $f(l,k) = \binom{l}{k} T_{l-k}$. □

Enumeration method for $f(l,k)$ (brute force). For each row, choose any q with $q \geq l+k$ (so every k-open length-l prefix has at least one extension to a length-q marked involution), enumerate the full set $\mathcal{S}_q$, group its elements by length-l prefix, restrict to groups whose prefix has exactly k open pairs at position l, and read off the number of such groups (= "distinct k-open prefixes") together with the maximum and minimum group size. The distinct count is independent of q (provided $q \geq l+k$), but the group size scales with q as $h(q-l,k)$. We vary the offset $q-(l+k) \in \{0,1,2\}$ across rows to demonstrate that the verification is not tied to the minimal q.

Exact-rational check. For each row we pick some $q \geq l+k$ (the slack $q-(l+k) \in \{0,1,2\}$ varies across rows), group $\mathcal{S}_q$ by length-l prefix, and restrict to groups whose prefix has k open pairs. We then record the number of such distinct k-open prefixes (the enumeration value of $f(l,k)$) together with the maximum and minimum group size over them. Two checks pass simultaneously: the count matches $f(l,k) = \binom{l}{k} T_{l-k}$ (independent of q once $q \geq l+k$), and max = min at every row – confirming the symmetry that every k-open length-l prefix has the same number of extensions in $\mathcal{S}_q$, namely $h(q-l,k)$ (varying with q).

(l,k)	offset	q	$\lvert\mathcal{S}_q\rvert = T_q$	distinct k-open	$f(l,k)$ closed	max	min	max = min
$(3,1)$	0	4	43	15	$\binom{3}{1}T_2 = 15$	1	1	✓
	1	5	142	15		4	4	✓
	2	6	499	15		15	15	✓
$(4,1)$	0	5	142	56	$\binom{4}{1}T_3 = 56$	1	1	✓
	1	6	499	56		4	4	✓
	2	7	1,850	56		15	15	✓
$(4,2)$	0	6	499	30	$\binom{4}{2}T_2 = 30$	2	2	✓
	1	7	1,850	30		12	12	✓
	2	8	7,193	30		60	60	✓
$(5,2)$	0	7	1,850	140	$\binom{5}{2}T_3 = 140$	2	2	✓
	1	8	7,193	140		12	12	✓
	2	9	29,186	140		60	60	✓
$(5,3)$	0	8	7,193	50	$\binom{5}{3}T_2 = 50$	6	6	✓
	1	9	29,186	50		48	48	✓
	2	10	123,109	50		300	300	✓
$(6,2)$	0	8	7,193	645	$\binom{6}{2}T_4 = 645$	2	2	✓
	1	9	29,186	645		12	12	✓
	2	10	123,109	645		60	60	✓
$(6,3)$	0	9	29,186	280	$\binom{6}{3}T_3 = 280$	6	6	✓
	1	10	123,109	280		48	48	✓
	2	11	538,078	280		300	300	✓
$(7,3)$	0	10	123,109	1,505	$\binom{7}{3}T_4 = 1{,}505$	6	6	✓
	1	11	538,078	1,505		48	48	✓
	2	12	2,430,355	1,505		300	300	✓
$(7,4)$	0	11	538,078	490	$\binom{7}{4}T_3 = 490$	24	24	✓
	1	12	2,430,355	490		240	240	✓
	2	13	11,317,646	490		1,800	1,800	✓
$(8,4)$	0	12	2,430,355	3,010	$\binom{8}{4}T_4 = 3{,}010$	24	24	✓
	1	13	11,317,646	3,010		240	240	✓
	2	14	54,229,907	3,010		1,800	1,800	✓

Table 1: Exact-rational verification of $f(l,k) = \binom{l}{k}T_{l-k}$ (Lemma 2). Each (l,k) pair is verified at three offsets $q-(l+k) \in \{0,1,2\}$. "distinct k-open" is the number of distinct length-l canonical prefixes appearing in $\mathcal{S}_q$ with k open pairs at position l, independent of the offset; "max" and "min" are the maximum/minimum of the group sizes $|\{\sigma \in \mathcal{S}_q : \sigma \text{ begins with } p\}|$ over those distinct k-open prefixes p. The fact that max = min = $h(q-l,k)$ at every row confirms the symmetry: every k-open length-l prefix has the same number of extensions in $\mathcal{S}_q$.

Lemma 3 (Closed form for h)**.** $h(r,k) = k!\binom{r}{k}T_{r-k}$.

Proof. Figure 2 illustrates the bijection.

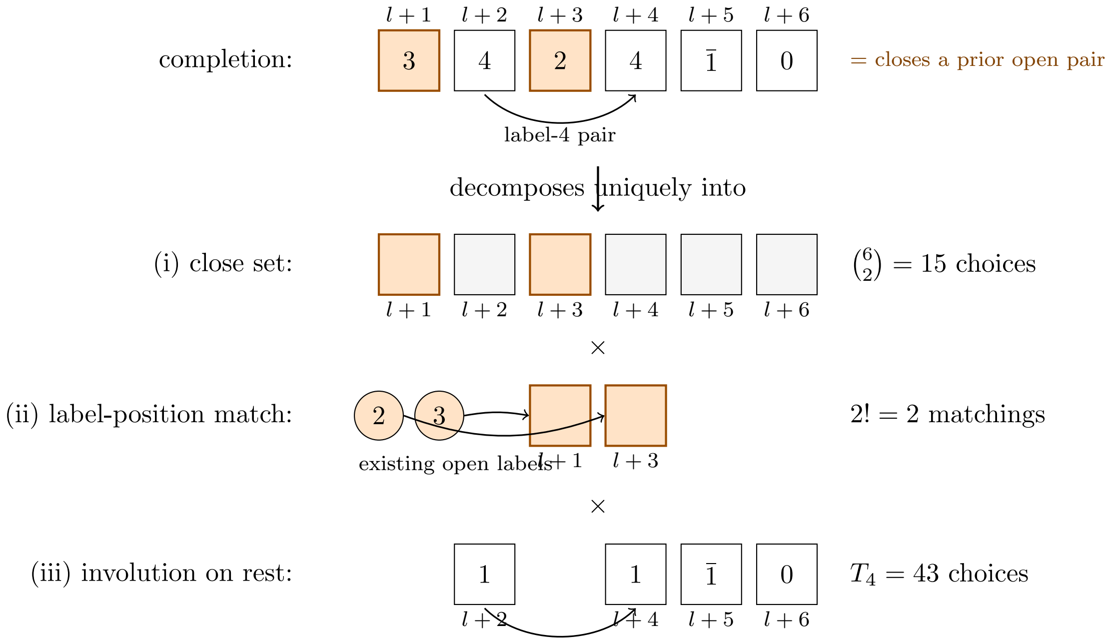

Figure 2: The bijection for Lemma 3: completing a k-open state over r further positions decomposes uniquely into (i) the set of k positions that close existing open pairs, (ii) a matching of the k pre-existing open labels to those positions, and (iii) a marked involution on the remaining $r-k$ positions. Example: $r=6$, $k=2$, continuing the prefix of Figure 1 (existing open labels $\{2,3\}$). Closing positions $\{l+1, l+3\}$ close labels 3 and 2 respectively; the remaining four positions $\{l+2, l+4, l+5, l+6\}$ carry the marked involution $l+2 \leftrightarrow l+4$, $l+5=\bar{1}$, $l+6=0$ (the new pair gets label 4 in the global encoding, the smallest integer not yet appearing). The full completion string is $3,4,2,4,\bar{1},0$. Total: $\binom{6}{2} \cdot 2! \cdot T_4 = 15 \cdot 2 \cdot 43 = 1{,}290 = h(6,2)$.

A completion of a k-open state over r further positions is determined by (i) the set of k positions among the r that close the existing open pairs ($\binom{r}{k}$ choices); (ii) the matching between the k pre-existing open labels and these k chosen positions ($k!$ ordered bijections); and (iii) a fresh marked involution on the remaining $r-k$ positions (T_{r-k} choices). Hence $h(r,k) = k!\binom{r}{k}T_{r-k}$. □

Enumeration method for $h(r,k)$ (brute force). For each row pick some $q \geq k+r$ (the slack $q-(k+r) \in \{0,1,2\}$ varies across rows), set $l = q-r$ (so $l \in \{k, k+1, k+2\}$), enumerate the full set $\mathcal{S}_q$, group by length-l prefix, restrict to groups whose prefix has k open pairs. The number of such groups equals $f(l,k)$ (varies with l; already verified in Table 1); the maximum and minimum group sizes are both $h(r,k)$, since every k-open length-l prefix has exactly $h(r,k)$ completions in $\mathcal{S}_{l+r}$ by symmetry. The offset-0 case ($l=k$, $f(l,k)=1$) is the original single-prefix enumeration; the offset-1, 2 cases give the same $h(r,k)$ over $f(l,k) > 1$ distinct prefixes.

Exact-rational check. For each row of Table 2 we group $\mathcal{S}_q$ (with $q = l+r$) by length-l prefix and restrict to k-open groups; the number of such groups equals $f(l,k)$ (column "distinct k-open"), and the maximum and minimum group size are recorded. Two checks pass simultaneously: max = min at every row (the symmetry that all k-open length-l prefixes have the same number of completions), and the common value equals $h(r,k) = k!\binom{r}{k}T_{r-k}$ independently of the choice of l within $\{k, k+1, k+2\}$ (the $h(r,k)$ value is shared down each (r,k) block of three rows).

(r,k)	offset	l	q=l+r	$\lvert\mathcal{S}_q\rvert$=$T_q$	distinct k-open	max	min	max = min =$h(r,k)$
$(3,1)$	0	1	4	43	1	15	15	$h(3,1)$=15
	1	2	5	142	4	15	15	
	2	3	6	499	15	15	15	
$(4,1)$	0	1	5	142	1	56	56	$h(4,1)$=56
	1	2	6	499	4	56	56	
	2	3	7	1,850	15	56	56	
$(4,2)$	0	2	6	499	1	60	60	$h(4,2)$=60
	1	3	7	1,850	6	60	60	
	2	4	8	7,193	30	60	60	
$(5,2)$	0	2	7	1,850	1	280	280	$h(5,2)$=280
	1	3	8	7,193	6	280	280	
	2	4	9	29,186	30	280	280	
$(5,3)$	0	3	8	7,193	1	300	300	$h(5,3)$=300
	1	4	9	29,186	8	300	300	
	2	5	10	123,109	50	300	300	
$(6,2)$	0	2	8	7,193	1	1,290	1,290	$h(6,2)$=1,290
	1	3	9	29,186	6	1,290	1,290	
	2	4	10	123,109	30	1,290	1,290	
$(6,3)$	0	3	9	29,186	1	1,680	1,680	$h(6,3)$=1,680
	1	4	10	123,109	8	1,680	1,680	
	2	5	11	538,078	50	1,680	1,680	
$(7,3)$	0	3	10	123,109	1	9,030	9,030	$h(7,3)$=9,030
	1	4	11	538,078	8	9,030	9,030	
	2	5	12	2,430,355	50	9,030	9,030	
$(7,4)$	0	4	11	538,078	1	11,760	11,760	$h(7,4)$=11,760
	1	5	12	2,430,355	10	11,760	11,760	
	2	6	13	11,317,646	75	11,760	11,760	
$(8,4)$	0	4	12	2,430,355	1	72,240	72,240	$h(8,4)$=72,240
	1	5	13	11,317,646	10	72,240	72,240	
	2	6	14	54,229,907	75	72,240	72,240	

Table 2: Exact-rational verification of $h(r,k) = k!\binom{r}{k}T_{r-k}$ (Lemma 3). Each (r,k) pair is verified at three offsets $q-(k+r) \in \{0,1,2\}$, giving $l = q-r \in \{k,k+1,k+2\}$. "distinct k-open" is the number of distinct length-l canonical k-open prefixes (equals $f(l,k)$; varies with l). The fact that max = min at every row, all three equal to $h(r,k)$ independently of l, confirms the symmetry: every k-open length-l prefix has the same number of completions in $\mathcal{S}_{l+r}$, namely $h(r,k)$.

Lemma 4 (Keystone identity)**.** *For* $0 \le l \le q$*,*

$$\sum_{k=0}^{l} f(l,k)\, h(q-l,k) \;=\; T_q. \tag{3}$$

Proof. Figure 3 illustrates the partition.

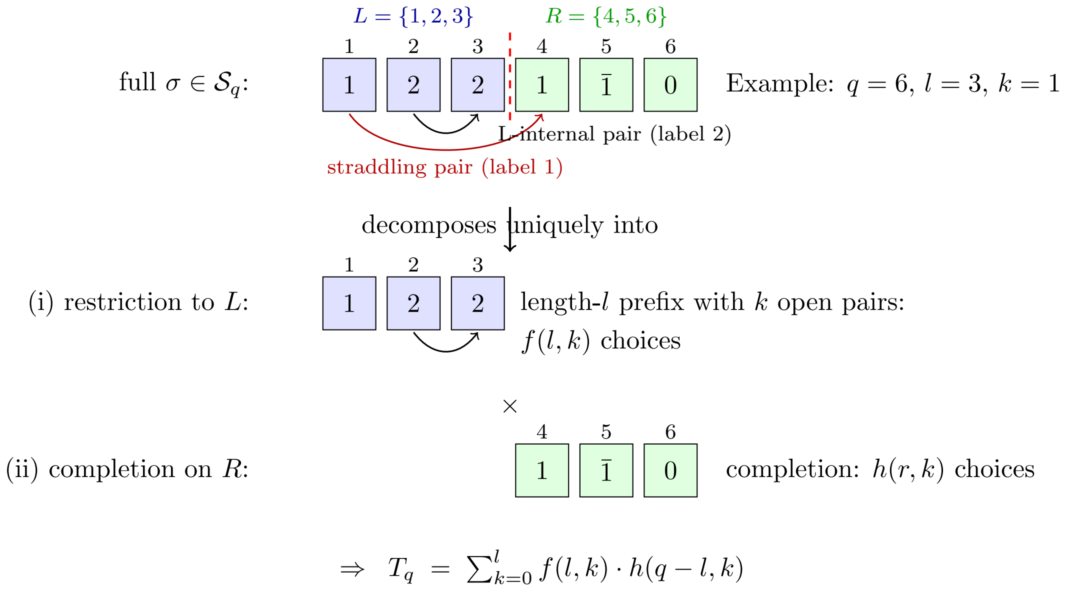

Figure 3: The partition for Lemma 4: a length-q marked involution decomposes by the number k of pairs that straddle the boundary $L \cup R$ ($|L| = l$, $|R| = r = q - l$), into (i) a length-l prefix with k open pairs (contributing $f(l,k)$ choices for each k) and (ii) a completion of that k-open state over r further positions (contributing $h(r,k)$). Example shown: $\sigma = 1\,2\,2\,1\,\bar{1}\,0$ in $\mathcal{S}_6$ with $l = 3$, $k = 1$ (one straddling pair labeled 1, one L-internal closed pair labeled 2, two R-internal fixed points). Summing over k gives T_q.

Setting $r := q - l$, we partition marked involutions of order $l + r$ by the number k of pairs that *straddle* the boundary between $L := \{1, \ldots, l\}$ and $R := \{l+1, \ldots, l+r\}$ (i.e. have one endpoint in L and the other in R).

A marked involution with exactly k straddling pairs is determined by two independent data:

- Its restriction to L, which is a length-l prefix in which the k left endpoints of the straddling pairs are precisely the positions that are *still open* at the end of the prefix; this contributes $f(l,k)$ choices.
- A completion of that k-open state into a full marked involution by filling the r further positions of R: this is counted by $h(r,k)$.

Summing over k gives $\sum_{k=0}^{l} f(l,k)h(r,k) = T_q$. □

Exact-rational check. For every $q \in \{2,3,4,5,6\}$ and every $l \in \{0,1,\ldots,q\}$ we compute the sum $\sum_{k=0}^{l} f(l,k)h(q-l,k)$ from the closed forms of Lemmas 2 and 3 and compare with T_q; all $3+4+5+6+7 = 25$ entries match (Table 3 below).

q	l	$\sum_{k=0}^{l} f(l,k)h(q-l,k)$	T_q	match
2	0	5	5	✓
	1	5	5	✓
	2	5	5	✓
3	0	14	14	✓
	1	14	14	✓
	2	14	14	✓
	3	14	14	✓
4	0	43	43	✓
	1	43	43	✓
	2	43	43	✓
	3	43	43	✓
	4	43	43	✓
5	0	142	142	✓
	1	142	142	✓
	2	142	142	✓
	3	142	142	✓
	4	142	142	✓
	5	142	142	✓
6	0	499	499	✓
	1	499	499	✓
	2	499	499	✓
	3	499	499	✓
	4	499	499	✓
	5	499	499	✓
	6	499	499	✓

Table 3: Exact-rational verification of the keystone identity $\sum_{k=0}^{l} f(l,k)h(q-l,k) = T_q$ (Lemma 4). All (q,l) with $2 \le q \le 6$ and $0 \le l \le q$ are checked exhaustively; all 25 entries match.

Substituting the closed forms of Lemmas 2 and 3 into (3) yields the equivalent Vandermonde-style algebraic identity

$$T_{l+r} \;=\; \sum_{k\geq 0} \binom{l}{k}\binom{r}{k} k!\, T_{l-k} T_{r-k}, \tag{4}$$

a classical identity for the EGF $e^{2z+z^2/2}$ recorded as an exercise in Riordan [8].

For a fixed length-l prefix p of order q with k_p open pairs, the probability that a uniformly random marked involution of order q begins with p is $\pi_p := h(q-l,k_p)/T_q$.

The trie-node and pointer counts of §4 are averages over n i.i.d. keys, so each reduces by inclusion–exclusion on the event "prefix p appears in the trie" = "at least one of the n keys begins with p". The indicator of this event has expectation $1-(1-\pi_p)^n = \sum_{j\geq 1}(-1)^{j+1}\binom{n}{j}\pi_p^j$, so its j-th term contributes the sum $\sum_p \pi_p^j$ over prefixes. The node count is just $\sum_p \mathbf{1}[p \text{ in trie}]$, so it reduces to the *unweighted* moments $\sum_p \pi_p^j$. The pointer count is $\sum_p d_p\, \mathbf{1}[p \text{ in trie}]$ with $d_p = k_p + 3$ by (1) except on two thin boundary lines where it drops by 1 or 3, so it reduces to a linear combination of the unweighted moments and the *TMS-weighted* moments $\sum_p k_p\, \pi_p^j$. We therefore define

$$S_j(q) \;:=\; 1 + \sum_{l=1}^{q} \sum_{p:|p|=l} \pi_p^j \;=\; 1 + \sum_{l=1}^{q}\sum_{k=0}^{l} f(l,k)\Big(\frac{h(q-l,k)}{T_q}\Big)^j, \tag{5}$$

$$S_j^{(k)}(q) \;:=\; \sum_{l=1}^{q} \sum_{p:|p|=l} k_p\, \pi_p^j \;=\; \sum_{l=1}^{q}\sum_{k=0}^{l} k\, f(l,k)\Big(\frac{h(q-l,k)}{T_q}\Big)^j. \tag{6}$$

The leading "1" in S_j is the contribution of the root prefix $p = \varepsilon$ at level $l = 0$, which has $\pi_\varepsilon = 1$ and so contributes $1^j = 1$ for every j; the same root contributes $k_\varepsilon \cdot 1 = 0$ to $S_j^{(k)}$ (no open pairs at the root), which is why no such "1+" appears there.

The second equality in each of (5) and (6) deserves a quick justification: $\pi_p = h(q - |p|, k_p)/T_q$ depends on the prefix p only through its length $|p|$ and its open-pair count k_p. So if we group the length-l prefixes by their open-pair count, the number of p with $|p| = l$ and $k_p = k$ is exactly $f(l,k)$ (Lemma 2), each contributing $\pi_p^j = (h(q-l,k)/T_q)^j$ (respectively $k_p\,\pi_p^j = k\,(h(q-l,k)/T_q)^j$ for $S_j^{(k)}$). Replacing $\sum_{p:|p|=l}$ by $\sum_{k=0}^{l} f(l,k)$ gives the two (l,k)-cell forms.

At $j = 1$, $\sum_{p:|p|=l} \pi_p = 1$ at every level l by (3), so $S_1(q) = 1 + q$.

Lemma 5 ($S_1^{(k)}$ in closed form)**.** *For all $q \geq 2$,*

$$S_1^{(k)}(q) \;=\; \frac{T_{q-2}}{T_q} \cdot \frac{q^3 - q}{6}. \tag{7}$$

(For $q \in \{0, 1\}$ the sum is empty, $S_1^{(k)}(q) = 0$.)

Proof. At fixed level l, using $k\binom{l}{k} = l\binom{l-1}{k-1}$ and the matching identity $(m+1)!\binom{q-l}{m+1} = (q - l)\,m!\binom{q-l-1}{m}$ on the h-factor, then shifting the summation index $k \to m+1$,

$$\begin{aligned}
\sum_k k\,f(l,k)\,h(q-l,k) &= \sum_{k\geq 1} k\binom{l}{k} T_{l-k} \cdot k!\binom{q-l}{k} T_{q-l-k} \\
&= \sum_{m\geq 0} l\binom{l-1}{m} T_{l-1-m} \cdot (q-l)\,m!\binom{q-l-1}{m} T_{q-l-1-m} \\
&= l(q-l)\sum_{m\geq 0}\binom{l-1}{m}\binom{q-l-1}{m} m!\,T_{l-1-m}T_{q-l-1-m} \\
&= l(q-l)\,T_{q-2}
\end{aligned}$$

by the keystone identity (3) applied at total length $(l-1) + (q-l-1) = q-2$. Summing over $l = 1, \ldots, q$ and dividing by T_q gives (7), since $\sum_{l=1}^{q} l(q-l) = (q^3 - q)/6$. □

Exact-rational check. For each $q \in \{3, 4, \ldots, 10\}$ we compute $S_1^{(k)}(q)$ via two *separate closed-form expressions* and check agreement: (i) the definition unwound via Lemmas 2 and 3,

$$S_1^{(k)}(q) \;=\; \sum_{l=1}^{q}\sum_{k=0}^{l} k\,f(l,k)\,h(q-l,k)/T_q,$$

evaluated as a `BigRational`; (ii) the Lemma 5 closed form $T_{q-2}/T_q \cdot (q^3 - q)/6$. Both are pure formula evaluations; this is not a brute-force enumeration check but an algebraic agreement test.

q	by definition (i)	by Lemma 5 (ii)	match
3	4/7	4/7	✓
4	50/43	50/43	✓
5	140/71	140/71	✓
6	1505/499	1505/499	✓
7	3976/925	3976/925	✓
8	41,916/7,193	41,916/7,193	✓
9	111,000/14,593	111,000/14,593	✓
10	1,186,845/123,109	1,186,845/123,109	✓

Table 4: Verification of Lemma 5 by computing $S_1^{(k)}(q)$ via two separate closed-form expressions: (i) the definition $\sum_{l,k} k\, f(l,k)\, h(q-l,k)/T_q$ from Lemmas 2, 3, and (ii) the Lemma 5 closed form $T_{q-2}/T_q \cdot (q^3-q)/6$. The two formulas produce the same `BigRational` at every q tested.

4 Reductions

Two reductions, one for the node count, one for the pointer count.

Theorem 6 (Reduction for P_{nodes})**.** *For all integers $n \geq 1$ and $q \geq 0$,*

$$P_{nodes}(n,q) \; = \; nq + (n-1) \; - \; \sum_{j=2}^{n} (-1)^j \binom{n}{j} S_j(q). \tag{8}$$

Proof. By definition $P_{\text{nodes}}(n,q) = \sum_{l=1}^{q} \sum_{p:|p|=l} \big(1-(1-\pi_p)^n\big)$, the inner sum running over the (finitely many) length-l prefixes of order-q marked involutions and $\big(1-(1-\pi_p)^n\big)$ being the probability that at least one of the n random keys begins with p. Expanding $1-(1-x)^n = \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}x^j$ in the coordinate $x = \pi_p$ and exchanging summation order,

$$P_{\text{nodes}}(n,q) = \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}\sum_{l=1}^{q}\sum_{p:|p|=l}\pi_p^j = \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}(S_j(q)-1).$$

For $j=1$, $S_1(q)-1 = q$ by (3); this contributes nq. The constant offset coming from $-\sum_{j\geq 2}(-1)^{j+1}\binom{n}{j}\cdot 1 = \sum_{j=2}^{n}(-1)^j\binom{n}{j} = n-1$ (since $\sum_{j=0}^{n}(-1)^j\binom{n}{j} = 0$ for $n \geq 1$, so the $j \geq 2$ tail equals $-[(-1)^0\binom{n}{0} + (-1)^1\binom{n}{1}] = n-1$). □

Exact-rational check. We verify (8) at small (n,q) by two independent computations: (i) "enumeration": enumerate all T_q^n ordered n-tuples of length-q marked involutions, build the compact trie of each tuple, count its distinct non-root nodes, and average as a `BigRational` (this code path uses no f, h formulas, only the trie construction); (ii) "closed form": compute $S_j(q)$ for $j = 2, \ldots, n$ from $S_j(q) = 1 + \sum_{(l,k)} f(l,k)(h(q-l,k)/T_q)^j$ using Lemmas 2 and 3, then plug into $P_{\text{nodes}}(n,q) = nq + (n-1) - \sum_{j=2}^{n}(-1)^j\binom{n}{j}S_j(q)$. If Theorem 6 holds, both columns must agree digit for digit; Table 5 confirms this.

(n,q)	T_q^n tuples	enumeration	closed form	match
$(2,4)$	1,849	13868/1849	13868/1849	✓
$(2,5)$	20,164	95899/10082	95899/10082	✓
$(2,6)$	249,001	2865678/249001	2865678/249001	✓
$(3,4)$	79,507	844810/79507	844810/79507	✓
$(3,5)$	2,863,288	19561753/1431644	19561753/1431644	✓
$(3,6)$	124,251,499	2070484608/124251499	2070484608/124251499	✓
$(4,4)$	3,418,801	46024640/3418801	46024640/3418801	✓
$(4,5)$	406,586,896	3566149555/203293448	3566149555/203293448	✓

Table 5: Verification of Theorem 6 by two independent computations of $P_{\text{nodes}}(n,q)$: "enumeration" (build the compact trie of every ordered n-tuple, count nodes, average) and "closed form" (apply (8) with $S_j(q)$ computed from f, h). Both columns match digit for digit at every tested (n,q).

Theorem 7 (Reduction for P_{ptr})**.** *For all integers $n \geq 1$ and $q \geq 2$ (so that the root's $l = k = 0$ satisfies $l + k = 0 \leq q - 2$, placing it in the "otherwise" branch of (1) with $d(0,0) = 0 + 3 = 3$),*

$$P_{ptr}(n,q) \;=\; 3P_{nodes}(n,q) \;+\; 3 \;+\; W(n,q) \;+\; R(n,q), \tag{9}$$

where the additive $+3$ is the root contribution ($d_{\text{root}} = 3$, not counted in P_{nodes}), and

$$W(n,q) \;=\; \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} S_j^{(k)}(q), \qquad R(n,q) \;=\; \mathbf{E}\Big[\sum_{p \in \text{Trie}} r_p\Big],$$

with $r_p \in \{-3, -1, 0\}$ the per-node correction $d_p - (k_p + 3)$ from (1) (so $r_p = -1$ at the "tight $= q-1$" boundary $l_p + k_p = q - 1$, $r_p = -3$ at the "tight $= q$" boundary $l_p + k_p = q$, and $r_p = 0$ elsewhere). The two boundary lines $l_p + k_p \in \{q-1, q\}$ each carry $\Theta(q)$ candidate cells, but the probability-weighted contribution $\mathbf{E}\big[\sum_p \mathbf{1}[p \text{ on boundary, in trie}]\big]$ scales like $\sqrt{q}$ (a $1/\sqrt{q}$ saturation density times $\Theta(q)$ length), so $R(n,q)$ is not *a bounded* $\mathrm{O}(1)$ *quantity; in fact $R(1,q) = (c_1 - 2)\sqrt{q} + \mathrm{O}(1)$ with $c_1 := \int_0^\infty e^{-\xi^2/2 - 2\xi} d\xi \approx 0.42136923$ (§9 below), and $R(n,q) = n\, R(1,q)$ up to an exponentially small remainder $\mathrm{O}(e^{-q \log 2\,(1+o(1))})$, by Lemma 17.*

Proof. The pointer count sums d_p over internal nodes p of the trie (levels $0 \leq l_p \leq q-1$); the leaf level $l_p = q$ holds the terminal involution data (the "*WordTable*" of exercise 204) and contributes no d. Equivalently, since $d(q,0) = 0$ in the tight branch of (1), we may include the leaf level without changing the value, and separate out the root:

$$P_{\text{ptr}}(n,q) \;=\; \mathbf{E}\Big[\sum_{p \in \text{Trie}} d_p\Big] \;=\; d(0,0) \;+\; \mathbf{E}\Big[\sum_{\substack{p \in \text{Trie} \\ |p| \geq 1}} d_p\Big].$$

The root is in the trie deterministically, so the second sum collapses to one over (l,k) cells with $1 \leq l \leq q$, $0 \leq k \leq l$. Each such cell contains $f(l,k)$ candidate prefixes, all sharing the same probability $\pi_{l,k} := h(q-l,k)/T_q$ that a random key begins with any one of them; the probability that the trie contains a given such prefix is $1 - (1 - \pi_{l,k})^n$. Hence

$$P_{\text{ptr}}(n,q) \;=\; 3 \;+\; \sum_{l=1}^{q} \sum_{k=0}^{l} d(l,k)\, f(l,k)\big(1 - (1 - \pi_{l,k})^n\big),$$

using $d_{\text{root}} = d(0,0) = 3$.

We decompose $d(l,k)$ into a baseline plus two boundary corrections:

$$d(l,k) \;=\; (k+3) \;-\; \mathbf{1}[l + k = q - 1] \;-\; 3\,\mathbf{1}[l + k = q], \tag{10}$$

which one checks against (1) case by case (at $l+k=q$: $k+3-0-3=k$; at $l+k=q-1$: $k+3-1-0=k+2$; otherwise: $k+3-0-0=k+3$). Substituting (10) into the P_{ptr}-expression splits the double sum into three pieces:

$$\begin{aligned}P_{\mathrm{ptr}}(n,q)-3 \;=\; & \sum_{l,k}(k+3)\,f(l,k)\big(1-(1-\pi_{l,k})^n\big) \\ & -\;\sum_{l,k}\mathbf{1}[l+k=q-1]\,f(l,k)\big(1-(1-\pi_{l,k})^n\big) \\ & -\;3\sum_{l,k}\mathbf{1}[l+k=q]\,f(l,k)\big(1-(1-\pi_{l,k})^n\big).\end{aligned}$$

Expand each $(1-(1-\pi_{l,k})^n)$ via the binomial identity $1-(1-x)^n=\sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}x^j$ and exchange the j-sum with the (l,k)-sums. For the first piece, the inner j-th term is

$$\sum_{l=1}^{q}\sum_{k=0}^{l}(k+3)\,f(l,k)\,\pi_{l,k}^{j} \;=\; \underbrace{\sum_{l,k}k\,f(l,k)\pi_{l,k}^{j}}_{=S_j^{(k)}(q)} + 3\underbrace{\sum_{l,k}f(l,k)\pi_{l,k}^{j}}_{=S_j(q)-1},$$

by (5) and (6) (the "-1" is the root term $\pi_{0,0}^{j}=1$ in S_j). For the second and third pieces, define the *boundary moments*

$$B_j^{(b)}(q) \;:=\; \sum_{(l,k):\,q-l-k=b} f(l,k)\,\pi_{l,k}^{j} \qquad (b=0,1), \tag{11}$$

so the inner j-th terms are $B_j^{(1)}(q)$ and $B_j^{(0)}(q)$ respectively. Combining,

$$P_{\mathrm{ptr}}(n,q) \;=\; 3 \;+\; \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}\big[S_j^{(k)}(q)+3\big(S_j(q)-1\big)-B_j^{(1)}(q)-3B_j^{(0)}(q)\big].$$

The $3(S_j-1)$ sum reduces to $3P_{\mathrm{nodes}}(n,q)$ by Theorem 6; the $S_j^{(k)}$ sum is $W(n,q)$ by definition; the $B_j^{(b)}$ sums assemble into $-R(n,q)$ via $R(n,q):=-\sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}\big[B_j^{(1)}(q)+3B_j^{(0)}(q)\big]$, which is the inclusion–exclusion expansion of $\mathbf{E}[\sum_p r_p]$ with $r_p=d_p-(k_p+3)$ from (10). This yields (9). □

Exact-rational check. We verify (9) at small (n,q) by two independent computations: (i) "enumeration": enumerate all T_q^n ordered n-tuples, build the compact trie of each, sum the d_p rule (1) over distinct nodes, and average as a `BigRational` (this code path uses no f,h formulas, only the trie construction and the d_p rule); (ii) "closed form": for each (l,k) cell compute $d(l,k)\,f(l,k)\,(1-(1-\pi_{l,k})^n)$ from Lemmas 2, 3 with $\pi_{l,k}=h(q-l,k)/T_q$, plus the root term $d(0,0)=3$, and sum. If Theorem 7 holds (which is equivalent to the displayed sum (9) after the d-decomposition (10)), both columns must agree digit for digit; Table 6 confirms this.

(n,q)	T_q^n tuples	enumeration	closed form	match
$(1,3)$	14	54/7	54/7	✓
$(1,4)$	43	471/43	471/43	✓
$(1,6)$	499	9064/499	9064/499	✓
$(2,5)$	20,164	122042/5041	122042/5041	✓
$(3,4)$	79,507	1810095/79507	1810095/79507	✓
$(3,5)$	2,863,288	23488091/715822	23488091/715822	✓
$(4,4)$	3,418,801	93409173/3418801	93409173/3418801	✓

Table 6: Verification of Theorem 7 by two independent computations of $P_{\text{ptr}}(n,q)$: "enumeration" (build the compact trie of every ordered n-tuple, sum the d_p rule over distinct nodes, average) and "closed form" (sum $d(l,k)f(l,k)(1-(1-h(q-l,k)/T_q)^n)$ over (l,k) cells using Lemmas 2, 3). Both columns match digit for digit at every tested (n,q).

Proposition 8 (Closed form of $R(1,q)$)**.** *For $q \geq 2$,*

$$R(1,q) \;=\; -\sum_{c=0}^{\lfloor (q-1)/2\rfloor} \binom{q-1-c}{c} T_{q-1-2c}\,\frac{2(c+1)!}{T_q} \;-\; 3\sum_{c=0}^{\lfloor q/2\rfloor} \binom{q-c}{c} T_{q-2c}\,\frac{c!}{T_q}.$$

The first sum tracks the "$\mathrm{TMS}_l = q-1-l$" boundary line ($r_p=-1$); the second tracks the "$\mathrm{TMS}_l = q-l$" boundary ($r_p=-3$). Each term is a probability that a random length-q marked involution hits the corresponding boundary at some level l.

Proof. For $n=1$ the trie is a path: a length-l prefix p belongs to the trie precisely when it equals the length-l prefix of the single random key, which happens with probability $q_{l,k_p} = h(q-l,k_p)/T_q$. Linearity of expectation therefore gives, summing r_p over all (l,k) cells,

$$R(1,q) \;=\; \sum_{l,k} r_{l,k}\, f(l,k)\,\pi_{l,k} \;=\; -B_1^{(1)}(q) \;-\; 3\,B_1^{(0)}(q), \tag{12}$$

using $r_{l,k}=-1$ on the line $l+k=q-1$, $r_{l,k}=-3$ on $l+k=q$, and $r_{l,k}=0$ elsewhere (from (10)); the two boundary moments $B_1^{(b)}(q) = \sum_{(l,k):q-l-k=b} f(l,k)\pi_{l,k}$ are the $j=1$ specializations of (11). We now turn each boundary moment into the stated single sum.

First sum (line $l+k=q-1$, $r=-1$). On this line set $c := q-1-l$, so $k=c$ and $l=q-1-c$. The constraint $0 \le k \le l$ becomes $0 \le c \le \lfloor (q-1)/2 \rfloor$. The closed forms of Lemmas 2 and 3 give

$$f(l,k) \;=\; \binom{q-1-c}{c} T_{q-1-2c},$$

$$h(q-l,k) \;=\; h(c+1,c) \;=\; c!\binom{c+1}{c} T_1 \;=\; 2(c+1)!$$

(using $T_1=2$, the two markings of a single 1-cycle), so $f(l,k)\,\pi_{l,k} = \binom{q-1-c}{c} T_{q-1-2c}\cdot 2(c+1)!/T_q$ and

$$B_1^{(1)}(q) \;=\; \sum_{c=0}^{\lfloor (q-1)/2\rfloor} \binom{q-1-c}{c} T_{q-1-2c}\,\frac{2(c+1)!}{T_q}.$$

Second sum (line $l+k=q$, $r=-3$). Set $c := q-l$, so $k=c$ and $l=q-c$. The constraint $0 \le k \le l$ becomes $0 \le c \le \lfloor q/2 \rfloor$. Now

$$f(l,k) \;=\; \binom{q-c}{c} T_{q-2c}, \qquad h(q-l,k) \;=\; h(c,c) \;=\; c!\binom{c}{c} T_0 \;=\; c!,$$

so

$$B_1^{(0)}(q) \;=\; \sum_{c=0}^{\lfloor q/2\rfloor} \binom{q-c}{c} T_{q-2c}\,\frac{c!}{T_q}.$$

Substituting these two sums into (12) yields the stated formula. □

Exact-rational check. For each $q \in \{3, 4, 5, 6, 7\}$ we compute $R(1, q)$ by three independent methods and check they all agree:

(i) the direct definition (= Theorem 7 specialized to $n = 1$, expanded via Lemmas 2, 3):

$$R(1, q) \;=\; \sum_{(l,k):\, l+k \in \{q-1, q\}} r_{l,k}\, f(l, k)\, h(q - l, k)/T_q,$$

with $r_{l,k} = -1$ on $l + k = q - 1$ and $r_{l,k} = -3$ on $l + k = q$ (pure formula evaluation);

(ii) the Proposition 8 closed form (two single sums over c, pure formula evaluation);

(iii) true enumeration: for each $\sigma \in \mathcal{S}_q$, walk its canonical string, compute $\sum_{l=0}^{q} r(l, \mathrm{TMS}_l(\sigma))$ with r from (10), and average over σ as a `BigRational`. This method touches neither f nor h nor Proposition 8.

q	by definition (i)	by Prop. 8 (ii)	by enumeration (iii)	match
3	$-34/7$	$-34/7$	$-34/7$	✓
4	$-224/43$	$-224/43$	$-224/43$	✓
5	$-394/71$	$-394/71$	$-394/71$	✓
6	$-2920/499$	$-2920/499$	$-2920/499$	✓
7	$-5678/925$	$-5678/925$	$-5678/925$	✓

Table 7: Verification of Proposition 8 by three independent computations of $R(1, q)$: (i) direct definition as a sum over boundary-line (l, k) cells; (ii) the Proposition 8 closed form as two sums over c; (iii) true enumeration – iterate over every $\sigma \in \mathcal{S}_q$, sum $r(l, \mathrm{TMS}_l(\sigma))$ over all $q+1$ positions, and average. All three columns produce the same `BigRational`.

5 Summary of Part I

The reductions of Part I express both $P_{\mathrm{nodes}}(n, q)$ and $P_{\mathrm{ptr}}(n, q)$ as finite, exact linear combinations of the per-prefix moments

$$S_j(q) \;=\; 1 + \sum_{l=1}^{q} \sum_{k=0}^{l} f(l, k) \Big(\tfrac{h(q-l,k)}{T_q}\Big)^j, \qquad S_j^{(k)}(q) \;=\; \sum_{l=1}^{q} \sum_{k=0}^{l} k\, f(l, k) \Big(\tfrac{h(q-l,k)}{T_q}\Big)^j,$$

the boundary moments

$$B_j^{(b)}(q) \;=\; \sum_{(l,k):\, q-l-k=b} f(l, k) \Big(\tfrac{h(q-l,k)}{T_q}\Big)^j \qquad (b = 0, 1),$$

and the keystone identity $\sum_k f(l, k) h(q - l, k) = T_q$. Specifically:

$$P_{\mathrm{nodes}}(n, q) \;=\; nq + (n - 1) \;-\; \sum_{j=2}^{n} (-1)^j \binom{n}{j} S_j(q), \tag{13}$$

$$P_{\mathrm{ptr}}(n, q) \;=\; 3P_{\mathrm{nodes}}(n, q) + 3 + W(n, q) + R(n, q), \tag{14}$$

with

$$W(n, q) = \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} S_j^{(k)}(q), \qquad R(n, q) = -\sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} \big[B_j^{(1)}(q) + 3B_j^{(0)}(q)\big].$$

To turn (13)–(14) into closed-form asymptotic expansions, Part II will establish exactly three asymptotic facts about the building blocks:

(A) *Leading constant for S_j* (Lemma 12): $S_j(q) = c_j\sqrt{q} + \mathrm{const}_j + \mathrm{O}(q^{-1/2})$ for every fixed $j \geq 2$, where $c_j = \int_0^\infty \exp(-\frac{j}{2}\xi^2 - 2j\xi)\,d\xi$, $\mathrm{const}_j = \frac{1}{2} + \frac{j-2}{2}(\frac{1}{j} - 2c_j)$ for $j \geq 3$, with the $j = 2$ case $\mathrm{const}_2 = \frac{3}{2} - 4c_2$ differing from the general formula's $j = 2$ specialisation (which would give $\frac{1}{2}$) by $1 - 4c_2$, arising from the off-diagonal $a = 1$ slice contribution (see Case $j = 2$ in the proof of Lemma 12).

(B) *Asymptotic of $\rho(q) := T_{q-1}/T_q$* (Lemma 14): bootstrapped from the recurrence $\rho(q)[2 + (q-1)\rho(q-1)] = 1$, used to develop T_{q-2}/T_q and hence $S_1^{(k)}(q) = (q^3 - q)\,T_{q-2}/(6T_q)$ to four orders.

(C) *Asymptotic of $B_j^{(b)}$ and $S_j^{(k)}$* (§9): $B_j^{(b)}(q)$ is exponentially small for $j \geq 2$, with rate $-(j\log 2)/2$ per unit q (so $R(n,q) = n\,R(1,q) + \mathrm{O}(e^{-q\log 2\cdot(1+\mathrm{o}(1))})$), and $S_j^{(k)}(q) = c_j^{(k)}q + d_j^{(k)}\sqrt{q} + \mathrm{O}(1)$ with explicit rational + c_j-linear coefficients.

The Wright saddle-point asymptotic for T_N (eq. 15 below) is the sole external analytical input.

Part II
Asymptotic analysis

The second part turns the exact reductions of Part I into closed-form asymptotic expansions as $q \to \infty$. The analytical input comprises Stirling's formula, the saddle-point asymptotic of T_N ((15), due to Hayman's admissibility method), and a boundary-layer expansion in the scaled variable $\xi := k/\sqrt{q}$. Special-function content enters through the family $c_j := \int_0^\infty e^{-(j/2)\xi^2 - 2j\xi}\,d\xi$.

6 Leading constant c_j for S_j

The analytical input is the marked-involution saddle-point asymptotic of the Hayman-admissibility framework [3]; equivalently, this is the specialisation to $H(z) = e^{2z+z^2/2}$ of the saddle-point method for H-admissible functions developed by Hayman [3] (see also [4, Thm. VIII.4, Example VIII.5], with fixed-point weight 2). The $\mathrm{O}(N^{-1/2})$ remainder in the form below is Wright's:

$$\log T_N \;=\; \frac{N}{2}\log N - \frac{N}{2} + 2\sqrt{N} - \frac{1}{4} + \log C + r(N), \tag{15}$$

for an explicit positive constant C, with remainder $r(N) \to 0$ as $N \to \infty$ and the uniform bound $|r(N)| \leq K_0$ for all $N \geq 1$ ($K_0 \leq 1$ suffices; explicit derivation in §A.4):

$$r(N) \;\to\; 0 \quad (N \to \infty), \qquad |r(N)| \leq K_0 \quad \text{for all } N \geq 1. \tag{16}$$

The $-\frac{1}{4} + \log C$ piece cancels in every ratio $h(q-l,k)/T_q$ we form, so only the structural shape of (15) enters.

An aside on the complementary error function. The constants c_j defined in the abstract,

$$c_j \;:=\; \int_0^\infty e^{-(j/2)\xi^2 - 2j\xi}\,d\xi,$$

admit an elementary closed form via the complementary error function $\mathrm{erfc}(x) := \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-u^2}\,du$: completing the square $\xi^2/2 + 2\xi = (\xi+2)^2/2 - 2$ and substituting $u = \sqrt{j/2}\,(\xi+2)$ gives

$$c_j \;=\; \sqrt{\pi/(2j)}\;e^{2j}\;\mathrm{erfc}(\sqrt{2j}), \qquad j \geq 1, \tag{17}$$

which we will use occasionally for numerical evaluation. Numerically (to six digits): $c_1 \approx 0.421369$, $c_2 \approx 0.226339$, $c_3 \approx 0.155304$, $c_4 \approx 0.118326$, $c_5 \approx 0.095609$, $c_6 \approx 0.080224$, $c_7 \approx 0.069111$, $c_8 \approx 0.060706$, $c_9 \approx 0.054126$, $c_{10} \approx 0.048834$.

The single analytical workhorse for both the $c_j\sqrt{q}$ asymptotic of S_j (Lemma 12 below) and the asymptotic of the boundary moments $B_1^{(b)}$ that drive Lemma 15 is the following uniform expansion of the all-open prefix probability $h(q-k,k)/T_q$.

Lemma 9 (Boundary-layer profile). *There exist absolute constants $C^* \le 4$ and $q^* = 20$ such that for all $q \ge q^*$ and all integers k with $0 \le k \le \sqrt{q\log q}$, setting $\xi := k/\sqrt{q}$:*

$$\left| \frac{h(q-k,k)}{T_q} - e^{-\xi^2/2-2\xi} \right| \le C^* e^{-\xi^2/2-2\xi} \cdot \frac{\xi^3+1}{\sqrt{q}}. \tag{18}$$

(The bound $C^* \le 4$ comes from $|P(\xi)| \le 2(1+\xi^3)$ (§A) plus the exponentiation step. With $X := P(\xi)/\sqrt{q} + \mathrm{O}((\xi^4+1)/q + q^{-1/2})$ from (59), on the boundary layer $|X| \le \mathrm{o}(1)$, so $|e^X - 1| \le |X| + |X|^2/2 + \cdots \le 2|X|$, giving $|h/T_q - \varphi(\xi)| = \varphi(\xi)|e^X-1| \le 2\varphi(\xi)|P(\xi)|/\sqrt{q} + \mathrm{O}(\ldots) \le 4\varphi(\xi)(1+\xi^3)/\sqrt{q}$. Sharper constants are possible but not needed; the value $C^* \le 4$ suffices below.) Outside the boundary layer, for $\sqrt{q\log q} < k \le q/2$, $h(q-k,k)/T_q \le e^{-\xi^2/2}$ uniformly (see Figure 4), and consequently $\sum_{k>\sqrt{q\log q}}^{\lfloor q/2\rfloor} h(q-k,k)/T_q \le 2/\sqrt{\log q}$.

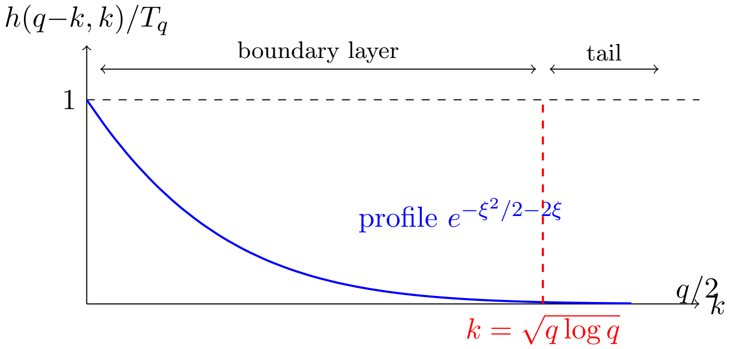

Figure 4: Decay regions in Lemma 9 (schematic; coefficients chosen to match $q = 16$ so that $1/(2q) \approx 0.031$ and $2/\sqrt{q} = 0.5$ both appear consistently). Inside the boundary layer ($k \le \sqrt{q\log q}$) the multiplicative profile $e^{-\xi^2/2-2\xi}$ approximates $h(q-k,k)/T_q$ to relative error $\mathrm{O}((\xi^3+1)/\sqrt{q})$. In the tail only the cruder upper bound $h(q-k,k)/T_q \le e^{-\xi^2/2}$ is needed, and the contribution of the entire tail is $\mathrm{O}(1/\sqrt{\log q})$.

Proof. By Stirling on factorials and Wright on T_q, T_{q-2k}, then Taylor expanding in $\xi = k/\sqrt{q}$, the leading $q\log q$, $\xi\sqrt{q}\log q$, $\xi\sqrt{q}$, $\frac{1}{2}\log q$ contributions all cancel (forced by $h(q-k,k)/T_q \le 1$). What survives is

$$\log\frac{h(q-k,k)}{T_q} = -\frac{\xi^2}{2} - 2\xi + \frac{P(\xi)}{\sqrt{q}} + \mathrm{O}\Big(\frac{\xi^4+1}{q}\Big) + \mathrm{O}(q^{-1/2}), \quad P(\xi) := -\frac{\xi^3}{2} - \xi^2 + \frac{\xi}{2}$$

Exponentiation on the boundary layer ($\xi \le \sqrt{\log q}$) yields (18). For the tail ($\sqrt{q\log q} < k \le q/2$): splitting at $\xi = q^{1/4}$, both the Taylor truncation regime ($\xi \le q^{1/4}$) and the non-truncated entropy regime ($\xi > q^{1/4}$, using $E(u) := (1-u)\log(1-u) - \frac{1-2u}{2}\log(1-2u) \le -u^2/2$ for $u \in [0,1/2)$, proved by series expansion in §A.2) give $h(q-k,k)/T_q \le e^{-\xi^2/2}$ uniformly; the boundary endpoint $k = \lfloor q/2\rfloor$ is $\le 2^{-q/2}e^{\mathrm{O}(\sqrt{q})}$ by direct Stirling+Wright. Riemann summation + Mills ratio yield $\sum_{\sqrt{q\log q}<k} h(q-k,k)/T_q \le 2/\sqrt{\log q}$.

Full Stirling+Wright bookkeeping for both regimes is in §A.1–§A.2. □

Table 8: Numerical verification of Lemma 9 (boundary-layer profile) at five representative k for each $q \in \{10, 100, 1000, 10000\}$. "Exact" is $h(q-k,k)/T_q = k!\binom{q-k}{k}T_{q-2k}/T_q$ in exact rational arithmetic; "Profile" is $e^{-\xi^2/2-2\xi}$ with $\xi = k/\sqrt{q}$; "Ratio" is Exact/Profile (should approach 1 on the boundary layer $\xi \le \sqrt{\log q}$, and diverges from 1 outside it since the true decay is faster than $e^{-\xi^2/2-2\xi}$). The middle column $k = \lfloor\sqrt{q\log q}\rfloor$ is the upper boundary of the layer where Lemma 9 guarantees $(1 + \mathrm{O}((\xi^3+1)/\sqrt{q}))$ accuracy. Entries written as "$< 10^{-300}$" indicate underflow below `float64` range; the corresponding Ratio is reported as "≈ 0" rather than computed.

q	k	ξ	Exact	Profile	Ratio
10	0	0.0000	1.0000	1.0000	1.0000
10	1	0.3162	5.2585×10^{-1}	5.0537×10^{-1}	1.0405
10	4	1.2649	1.4621×10^{-2}	3.5799×10^{-2}	0.4084
10	4	$(\lfloor\sqrt{q\log q}\rfloor = 4)$			
10	5	1.5811	9.7475×10^{-4}	1.2128×10^{-2}	0.0804
100	0	0.0000	1.0000	1.0000	1.0000
100	1	0.1000	8.1855×10^{-1}	8.1465×10^{-1}	1.0048
100	21	2.1000	6.0174×10^{-4}	1.6533×10^{-3}	0.3640
100	21	$(\lfloor\sqrt{q\log q}\rfloor = 21)$			
100	49	4.9000	1.4706×10^{-20}	3.3901×10^{-10}	≈ 0
100	50	5.0000	1.1534×10^{-22}	1.6919×10^{-10}	≈ 0
1000	0	0.0000	1.0000	1.0000	1.0000
1000	1	0.0316	9.3871×10^{-1}	9.3824×10^{-1}	1.0005
1000	83	2.6247	1.0114×10^{-4}	1.6761×10^{-4}	0.6034
1000	83	$(\lfloor\sqrt{q\log q}\rfloor = 83)$			
1000	499	15.78	$< 10^{-300}$	$< 10^{-50}$	≈ 0
1000	500	15.81	$< 10^{-300}$	$< 10^{-50}$	≈ 0
10000	0	0.0000	1.0000	1.0000	1.0000
10000	1	0.0100	9.8020×10^{-1}	9.8015×10^{-1}	1.0000
10000	303	3.0300	1.8959×10^{-5}	2.3690×10^{-5}	0.8003
10000	303	$(\lfloor\sqrt{q\log q}\rfloor = 303)$			
10000	4999	49.99	$< 10^{-300}$	$< 10^{-300}$	–
10000	5000	50.00	$< 10^{-300}$	$< 10^{-300}$	–

Reading the table. For each q, the ratio $\to 1$ as $k \to 0$ (small ξ, well inside the boundary layer where $(\xi^3+1)/\sqrt{q}$ is negligible), starts departing as ξ approaches $\sqrt{\log q}$ (the $\mathrm{O}((\xi^3+1)/\sqrt{q})$ correction becomes visible), and collapses far outside the layer as the true profile decays faster than $e^{-\xi^2/2-2\xi}$ (cf. the entropy-function bound in the "outside-the-boundary-layer" part of the proof of Lemma 9). For $k \in \{q/2-1, q/2\}$, both Exact and Profile are sub-poly small, so their numerical ratio is dominated by arithmetic precision; the multiplicative form $h(q-k,k)/T_q \le e^{-\xi^2/2}$ is what is actually used in Lemma 9's tail bound, and it holds with room to spare.

The diagonal asymptotic — a shared engine

Both S_j (§6, leading to P_{nodes}) and $S_j^{(k)}$ (§9, leading to $W(n,q)$ for $n \ge 2$) reduce to a sum over the diagonal $\{a = 0\}$ of the (a,k)-triangle, scaled by a power of k. We isolate the shared analytical work in one lemma below.

Lemma 10 (Diagonal asymptotic)**.** *For each fixed integer $j \ge 1$ and $\alpha \in \{0,1\}$,*

$$\sum_{k\ge1} k^\alpha \Big[\frac{h(q-k,k)}{T_q}\Big]^j \;=\; q^{(\alpha+1)/2}\, M_\alpha^{(j)} - \mathbf{1}[\alpha=0]\cdot\tfrac{1}{2} + q^{\alpha/2}\, j\int_0^\infty \xi^\alpha P(\xi) e^{-j\xi^2/2-2j\xi} d\xi + \mathrm{O}\big(q^{(\alpha-1)/2}\big), \tag{19}$$

where $P(\xi) := -\xi^3/2 - \xi^2 + \xi/2$ *is the profile correction from Lemma 9 and the moments are*

$$M_n^{(j)} := \int_0^\infty \xi^n e^{-j\xi^2/2-2j\xi} d\xi, \qquad M_0^{(j)} = c_j, \quad M_1^{(j)} = \tfrac{1}{j} - 2c_j.$$

Proof. Inside the boundary layer ($1 \le k \le \sqrt{q \log q}$), Lemma 9 gives, with $\xi = k/\sqrt{q}$,

$$[h(q-k,k)/T_q]^j \;=\; e^{-j\xi^2/2-2j\xi}\Big[1 + \frac{jP(\xi)}{\sqrt{q}} + \mathrm{O}\big((\xi^6+1)/q\big)\Big].$$

The contribution beyond the boundary layer is super-polynomially small (again Lemma 9, tail bound). Apply Euler–Maclaurin (trapezoidal rule with step $\Delta\xi = 1/\sqrt{q}$) to $\sum_{k\ge 0} k^\alpha e^{-j\xi_k^2/2-2j\xi_k}$. Since $k = \sqrt{q}\,\xi_k$, the integrand at $\xi = 0$ has value $\mathbf{1}[\alpha = 0]$:

$$\sum_{k\ge 0} k^\alpha e^{-j\xi_k^2/2-2j\xi_k} \;=\; q^{(\alpha+1)/2}\int_0^\infty \xi^\alpha e^{-j\xi^2/2-2j\xi} d\xi \;+\; \tfrac{1}{2}\,\mathbf{1}[\alpha=0] \;+\; \mathrm{O}(q^{(\alpha-1)/2}).$$

The first term is $q^{(\alpha+1)/2}\, M_\alpha^{(j)}$. The multiplicative $jP(\xi)/\sqrt{q}$ correction contributes

$$q^{\alpha/2}\, j \int_0^\infty \xi^\alpha P(\xi) e^{-j\xi^2/2-2j\xi}\, d\xi + \mathrm{O}(q^{(\alpha-1)/2})$$

(the trapezoidal endpoint at $\xi = 0$ vanishes since $P(0) = 0$).

The unified $\mathrm{O}(q^{(\alpha-1)/2})$ remainder absorbs three sources, each of that order: *(i)* the second-order Euler–Maclaurin endpoint term $(\Delta\xi/12)\cdot f'(0)\cdot q^{\alpha/2}$ where $\Delta\xi = 1/\sqrt{q}$; *(ii)* the boundary-layer-tail contribution $\sum_{k>\sqrt{q\log q}}$, which by Mills' ratio gives $\mathrm{O}(q^{\alpha/2}/\sqrt{\log q})$, in particular $\mathrm{O}(q^{(\alpha-1)/2})$ for any fixed power; and *(iii)* the $\mathrm{O}((\xi^3+1)/\sqrt{q})$ multiplicative profile error from Lemma 9, which contributes $q^{\alpha/2}\cdot \mathrm{O}(1/\sqrt{q}) = \mathrm{O}(q^{(\alpha-1)/2})$ after integration. Subtracting the $k = 0$ term (value $\mathbf{1}[\alpha = 0]$) yields (19). □

Remark 11. Specialisation to $\alpha = 0$ gives $\Sigma^{(0)}_{a=0} = c_j\sqrt{q} - \frac{1}{2} + j\int P\, e^{-j\xi^2/2-2j\xi} d\xi + \mathrm{o}(1)$ (used in Lemma 12). Specialisation to $\alpha = 1$ gives $\Sigma^{(1)}_{a=0} = qM_1^{(j)} + \sqrt{q}\, j\int \xi P\, e^{-j\xi^2/2-2j\xi} d\xi + \mathrm{o}(\sqrt{q})$ (used in Lemma 18). The two profile integrals evaluate via the moment recurrence $M_{n+1}^{(j)} = nM_{n-1}^{(j)}/j - 2M_n^{(j)}$ (using $[\xi^n e^{-j\xi^2/2-2j\xi}]_0^\infty = 0$ for $n \ge 1$):

$$j\int_0^\infty P(\xi)e^{-j\xi^2/2-2j\xi} d\xi = \tfrac{j-2}{2} M_1^{(j)},$$
$$j\int_0^\infty \xi P(\xi)e^{-j\xi^2/2-2j\xi} d\xi = \frac{(j-3)\big[(4j+1)c_j - 2\big]}{2j}.$$

Lemma 12 (Leading asymptotic of S_j)**.** *For each integer $j \ge 2$, as $q \to \infty$,*

$$\boxed{S_j(q) \;=\; c_j\sqrt{q} + \mathrm{const}_j + \mathrm{O}(q^{-1/2}), \qquad c_j \;:=\; \int_0^\infty e^{-(j/2)\xi^2-2j\xi}\, d\xi.} \tag{20}$$

Numerically, $c_1 \approx 0.42136923$*,* $c_2 \approx 0.22634$*,* $c_3 \approx 0.15530$*,* $c_4 \approx 0.11833$*; the constants* const_j *are also explicit:*

$$\mathrm{const}_2 \;=\; \tfrac{3}{2} - 4c_2 \;\approx\; 0.5946, \qquad \mathrm{const}_j \;=\; \tfrac{1}{2} \;+\; \tfrac{j-2}{2}\, M_1^{(j)} \;=\; \tfrac{1}{2} \;+\; \tfrac{j-2}{2}\Big(\tfrac{1}{j} - 2c_j\Big) \quad \textit{for } j \ge 3,$$

where $M_1^{(j)} := \int_0^\infty \xi\, e^{-(j/2)\xi^2-2j\xi} d\xi = 1/j - 2c_j$*. Numerically:* $\mathrm{const}_3 \approx 0.5114$, $\mathrm{const}_4 \approx 0.5133$, $\mathrm{const}_5 \approx 0.5132$*. The remainder beyond* $c_j\sqrt{q} + \mathrm{const}_j$ *is* $\mathrm{O}(q^{-1/2})$ *for all* $j \ge 2$*.*

Proof of Lemma 12. The proof reduces, via the boundary-layer profile (Lemma 9), to a Riemann sum plus an explicit accounting of off-diagonal slices.

Setup. By the second equality of (5), with the closed forms $f(l,k) = \binom{l}{k}T_{l-k}$ (Lemma 2) and $h(q-l,k) = k!\binom{q-l}{k}T_{q-l-k}$ (Lemma 3), reparametrise the (l,k) sum by $a := l-k \geq 0$ (so $l = a+k$, $q-l-k = q-a-2k$):

$$S_j(q) - 1 \;=\; \sum_{\substack{a,k\geq 0\\ a+2k\leq q\\ (a,k)\neq(0,0)}} \binom{a+k}{k} T_a \left[\frac{k!\binom{q-a-k}{k}T_{q-a-2k}}{T_q}\right]^j. \tag{21}$$

We split the sum by a: *diagonal* $(a=0)$, $k=0$ *column*, and *off-diagonal slices* $(a,k\geq 1)$.

Diagonal $a=0, k\geq 1$. On $a=0$, $\binom{k}{k}T_0 = 1$, and the summand is $[h(q-k,k)/T_q]^j$. Apply Lemma 10 with $\alpha = 0$ and substitute the profile integral evaluation $j\int P\, e^{-j\xi^2/2-2j\xi}d\xi = (j-2)M_1^{(j)}/2$ from Remark 11:

$$\Sigma_{a=0} \;:=\; \sum_{k\geq 1}\Big[\frac{h(q-k,k)}{T_q}\Big]^j \;=\; c_j\sqrt{q} - \tfrac{1}{2} + \tfrac{j-2}{2}M_1^{(j)} + \mathrm{o}(1), \tag{22}$$

which is $c_j\sqrt{q} - \frac{1}{2} + \mathrm{o}(1)$ for $j=2$ and $c_j\sqrt{q} - \frac{1}{2} + \frac{j-2}{2}(1/j - 2c_j) + \mathrm{o}(1)$ for $j\geq 3$.

$k=0$ column ($a\geq 1$). On $k=0$, $\binom{a}{0}T_a = T_a$ and $h(q-a,0)/T_q = T_{q-a}/T_q$. Directly from Wright's asymptotic (15), $\log(T_{q-a}/T_q) = \frac{q-a}{2}\log(q-a) - \frac{q}{2}\log q + (a/2) + 2(\sqrt{q-a}-\sqrt{q}) + \mathrm{O}(1) = -(a/2)\log q + \mathrm{O}(1)$, so $T_{q-a}/T_q = \mathrm{O}(q^{-a/2})$ uniformly for each fixed a. Hence

$$\Sigma_{k=0} \;:=\; \sum_{a\geq 1} T_a\big[T_{q-a}/T_q\big]^j \;=\; \mathrm{O}(q^{-j/2}),$$

since the $a=1$ term dominates ($T_1\cdot q^{-j/2} = 2q^{-j/2}$), $a=2$ contributes $\mathrm{O}(q^{-j})$, and higher a even less. So for $j\geq 2$, $\Sigma_{k=0} = \mathrm{o}(1)$.

Off-diagonal $a=1, k\geq 1$. With $\binom{1+k}{k}T_1 = 2(k+1)$, the $a=1$ summand is $2(k+1)\cdot[h(q-1-k,k)/T_q]^j$. Factor through size $q-1$: $h(q-1-k,k)/T_q = \rho(q)\cdot h(q-1-k,k)/T_{q-1}$, where the inner ratio is the size-$(q-1)$ profile (apply Lemma 9 at size $q-1$). Apply Lemma 10 at size $q-1$ with $\alpha=1$ then $\alpha=0$:

$$\sum_{k\geq 1}(k+1)\Big[\frac{h(q-1-k,k)}{T_{q-1}}\Big]^j \;=\; qM_1^{(j)} + c_j\sqrt{q} + \mathrm{O}(\sqrt{q} + q^{(1-j)/2}) \;=\; qM_1^{(j)} + \mathrm{O}(\sqrt{q}),$$

using $M_1^{(j)} = 1/j - 2c_j$ from Remark 11. Multiplying by $2\rho(q)^j = 2q^{-j/2}(1+\mathrm{O}(q^{-1/2}))$:

$$\Sigma_{a=1} \;=\; 2M_1^{(j)}\cdot q^{1-j/2} + \mathrm{O}(q^{(1-j)/2}). \tag{23}$$

For $j=2$: $\Sigma_{a=1} = 2M_1^{(2)} + \mathrm{o}(1) = 1 - 4c_2 + \mathrm{o}(1)$. For $j\geq 3$: $\Sigma_{a=1} = \mathrm{o}(1)$.

Off-diagonal $a\geq 2, k\geq 1$. By the same T_{q-a}/T_q factoring, $[h(q-a-k,k)/T_q]^j = \rho(q)^{aj}\cdot[h(q-a-k,k)/T_{q-a}]^j(1+o(1))$ on the boundary layer. So $\Sigma_{a\geq 2} = \mathrm{O}(q^{(a+1-aj)/2})$ summed over $a\geq 2$; the dominant $a=2$ term is $\mathrm{O}(q^{(3-2j)/2})$, which for $j\geq 2$ is $\mathrm{O}(q^{-1/2}) = \mathrm{o}(1)$.

Combining. $S_j(q) - 1 = \Sigma_{a=0} + \Sigma_{k=0} + \Sigma_{a=1} + \Sigma_{a\geq 2}$. With (22), $\Sigma_{k=0} = \mathrm{o}(1)$, $\Sigma_{a\geq 2} = \mathrm{o}(1)$, and (23):

- For $j\geq 3$: the $a=1$ slice is $\mathrm{o}(1)$, so

  $$S_j(q) = 1 + c_j\sqrt{q} - \tfrac{1}{2} + \tfrac{j-2}{2}M_1^{(j)} + \mathrm{o}(1) = c_j\sqrt{q} + \tfrac{1}{2} + \tfrac{j-2}{2}M_1^{(j)} + \mathrm{o}(1),$$

  i.e., $\mathrm{const}_j = \frac{1}{2} + \frac{j-2}{2}M_1^{(j)} = \frac{1}{2} + \frac{j-2}{2}(1/j - 2c_j)$.

- For $j = 2$: the diagonal correction vanishes ($(j-2)/2 = 0$), and the $a = 1$ slice contributes $1 - 4c_2$:

$$S_2(q) = 1 + c_2\sqrt{q} - \tfrac{1}{2} + 0 + (1 - 4c_2) + \mathrm{o}(1) = c_2\sqrt{q} + \tfrac{3}{2} - 4c_2 + \mathrm{o}(1),$$

  i.e., $\mathrm{const}_2 = \frac{3}{2} - 4c_2$.

This completes the proof. □

Table 9: Exact $S_j(q)$ vs. a leading two-term estimate "est" $= c_j\sqrt{q} + \mathrm{const}$, where const $= 3/2 - 4c_2$ for $j = 2$ (the full const_2 from Lemma 12) and const $= 1/2$ for $j \geq 3$ (the leading $\Sigma_{a=0} + (a,k) = (0,0)$ contribution alone, omitting the refinement $\frac{j-2}{2}M_1^{(j)}$ which is ≈ 0.011 for $j = 3$ and ≈ 0.013 for $j = 4$ — this refinement appears as the dominant $\mathrm{O}(1)$ part of the residual $S_j(q) - \mathrm{est}$). Also shown: the contributions of five disjoint cells of the (a,k)-triangle (which sum to $S_j(q) - 1$): $\Sigma_{a=0,1,2}$ are the rows $a \in \{0,1,2\}$ within $\{a + 2k \leq q/2\}$ (these are the dominant boundary-layer terms); $\Sigma_{a\geq 3}$ collects rows $a \geq 3$ within $\{a + 2k \leq q/2\}$; $\Sigma_{a+2k>q/2}$ collects everything in the upper half of the triangle (bulk + large-a tail). Computed in log-space float64 via the T-recurrence and Lemmas 2, 3; all 12 row sums verify $(S_j(q) - 1) = \Sigma_{a=0} + \Sigma_{a=1} + \Sigma_{a=2} + \Sigma_{a\geq 3} + \Sigma_{a+2k>q/2}$.

q	j	$S_j(q)$	Est	$\Sigma_{a=0}$	$\Sigma_{a=1}$	$\Sigma_{a=2}$	$\Sigma_{a\geq 3}$	$\Sigma_{a+2k>q/2}$
10	2	1.56091	1.31039	0.32804	0.18430	0.02914	6.74×10^{-3}	1.27×10^{-2}
10	3	1.19325	0.99111	0.15710	0.03428	0.00134	6.92×10^{-5}	4.56×10^{-4}
10	4	1.08640	0.87418	0.07912	0.00719	0.00007	8.54×10^{-7}	2.96×10^{-5}
100	2	2.93159	2.85803	1.80424	0.12079	0.00619	3.62×10^{-4}	2.64×10^{-9}
100	3	2.12538	2.05304	1.11894	0.00642	0.00002	9.64×10^{-8}	2.43×10^{-14}
100	4	1.76612	1.68326	0.76571	0.00041	0.00000	3.78×10^{-11}	4.95×10^{-19}
1000	2	7.77488	7.75210	6.67068	0.10257	0.00161	2.74×10^{-5}	~ 0
1000	3	5.44180	5.41114	4.44018	0.00162	0.00000	2.20×10^{-10}	~ 0
1000	4	4.27764	4.24180	3.27761	0.00003	0.00000	2.59×10^{-15}	~ 0
10000	2	23.23565	23.22850	22.13806	0.09711	0.00048	2.50×10^{-6}	~ 0
10000	3	16.04785	16.03039	15.04738	0.00047	0.00000	6.20×10^{-13}	~ 0
10000	4	12.35312	12.33262	11.35312	0.00000	0.00000	2.25×10^{-19}	~ 0
100000	2	72.17143	72.16917	71.07586	0.09542	0.00015	2.43×10^{-7}	~ 0
100000	3	49.62469	49.61140	48.62455	0.00015	0.00000	1.89×10^{-15}	~ 0
100000	4	37.93364	37.91803	36.93364	0.00000	0.00000	2.15×10^{-23}	~ 0

7 Combined: $P_{\mathbf{nodes}}$

Theorem 13 (P_{nodes}, four-term expansion)**.** *For every integer $n \geq 1$, as $q \to \infty$,*

$$\boxed{P_{nodes}(n,q) \;=\; nq \;+\; (n-1) \;-\; c_n^*\sqrt{q} \;-\; d_n^* \;+\; \mathrm{O}(q^{-1/2}),} \tag{24}$$

where

$$c_n^* \;:=\; \sum_{j=2}^{n}(-1)^j\binom{n}{j}c_j \;=\; \int_0^\infty \big[(1-\varphi(\xi))^n + n\,\varphi(\xi) - 1\big]d\xi, \tag{25}$$

$$d_n^* \;:=\; \sum_{j=2}^{n}(-1)^j\binom{n}{j}\,\mathrm{const}_j, \tag{26}$$

in perfect symmetry with c_n^**, so that the* $\sqrt{q}$ *and* $\mathrm{O}(1)$ *corrections combine to* $c_n^*\sqrt{q} + d_n^* = \sum_{j=2}^n (-1)^j \binom{n}{j}\big(c_j\sqrt{q} + \mathrm{const}_j\big)$*, with* $\varphi(\xi) := \exp(-\xi^2/2 - 2\xi)$ *and* c_j*,* const_j *from Lemma 12. The* $j = 2$ *case is anomalous: the off-diagonal* $a = 1$ *slice contributes at the same order and gives* $\mathrm{const}_2 = 3/2 - 4c_2$ *rather than the naive* $j = 2$ *value* $1/2$ *from the general formula. The* $\mathrm{O}(q^{-1/2})$ *remainder is inherited from the worst-case* ($j = 2$) *residual in Lemma 12; at* $n = 1$ *it vanishes identically because* $P_{nodes}(1, q) = q$ *exactly. The first two coefficients* nq *and* $(n-1)$ *are exact (from the keystone identity at every level and from* $\sum_{j=2}^n (-1)^j \binom{n}{j} = n - 1$*); the* $-c_n^*\sqrt{q}$ *correction reflects the boundary-layer prefix-coincidence effect; the constant* $-d_n^*$ *absorbs the sub-leading* const_j *contributions hidden inside the original* $\mathrm{O}(1)$ *remainder.*

Proof. Combine Theorem 6

$$P_{\text{nodes}}(n, q) = nq + (n-1) - \sum_{j=2}^{n} (-1)^j \binom{n}{j} S_j(q)$$

with the three-term expansion $S_j(q) = c_j\sqrt{q} + \mathrm{const}_j + R_j(q)$ (Lemma 12, where $R_j(q) = \mathrm{O}(q^{-1/2})$ for $j \in \{2, 3\}$ and $\mathrm{O}(q^{(2-j)/2})$ for $j \geq 4$; both $j = 2$ and $j = 3$ give the same $q^{-1/2}$ rate, $j \geq 4$ strictly faster). Substituting:

$$\begin{aligned} P_{\text{nodes}}(n, q) &= nq + (n-1) \\ &\quad - \sqrt{q}\sum_{j=2}^{n} (-1)^j \binom{n}{j} c_j \\ &\quad - \sum_{j=2}^{n} (-1)^j \binom{n}{j} \mathrm{const}_j \\ &\quad - \sum_{j=2}^{n} (-1)^j \binom{n}{j} R_j(q), \end{aligned}$$

which identifies the three coefficient sums as $-c_n^*\sqrt{q}$, $-d_n^*$, and $-R_n(q)$. The dominant residual comes from $j \in \{2, 3\}$ both at rate $\mathrm{O}(q^{-1/2})$ (with $j \geq 4$ contributing $\mathrm{O}(q^{(2-j)/2})$, strictly faster).

Integral form for c_n^**.* By Lemma 12, $c_j = \int_0^\infty \varphi(\xi)^j \, d\xi$ with $\varphi(\xi) = e^{-\xi^2/2 - 2\xi}$. Interchange sum and integral (the sum is finite, no convergence issue):

$$c_n^* \;=\; \sum_{j=2}^{n} (-1)^j \binom{n}{j} \int_0^\infty \varphi^j \, d\xi \;=\; \int_0^\infty \Bigg[\sum_{j=2}^{n} (-1)^j \binom{n}{j} \varphi^j\Bigg] d\xi.$$

The bracketed sum is the binomial $(1-\varphi)^n = \sum_{j=0}^n (-1)^j \binom{n}{j}\varphi^j$ with the $j = 0, 1$ terms removed:

$$\sum_{j=2}^{n} (-1)^j \binom{n}{j} \varphi^j \;=\; (1-\varphi)^n - \underbrace{\binom{n}{0}}_{=1} + \underbrace{\binom{n}{1}\varphi}_{=n\varphi} \;=\; (1-\varphi)^n + n\varphi - 1.$$

Substituting back yields the integral form $c_n^* = \int_0^\infty [(1-\varphi)^n + n\varphi - 1] \, d\xi$ recorded in (25). Split: $c_n^* = n\, c_1 + \int_0^\infty [(1-\varphi)^n - 1] \, d\xi$. On the second integrand, $(1-\varphi)^n \to 0$ when $n\varphi \gg 1$ and $(1-\varphi)^n \to 1$ when $n\varphi \ll 1$. The crossover happens at $\varphi(\xi^*) = (\log 2)/n$, i.e. at the solution of $\xi^{*2}/2 + 2\xi^* = \log n - \log\log 2$; since the quadratic term dominates for large n, $\xi^* = \sqrt{2\log n} - 2 + \mathrm{O}((\log n)^{-1/2})$. Hence

$$\int_0^\infty [(1-\varphi)^n - 1] \, d\xi \;=\; -\xi^* + \mathrm{O}(1) \;=\; -\sqrt{2\log n} + \mathrm{O}(1),$$

giving $c_n^* = n\,c_1 - \sqrt{2\log n} + \mathrm{O}(1)$ as $n \to \infty$ — c_n^* grows *linearly* in n with slope $c_1 \approx 0.42136923$, shifted down by a $\sqrt{2\log n}$ correction. Numerically at $n = 1000$: $1000\,c_1 - \sqrt{2\log 1000} \approx 417.65$ versus actual $c_{1000}^* \approx 419.02$ (consistent with an $\mathrm{O}(1)$ residual ≈ 1.4).

Derivation of the closed form (26). By Lemma 12, $\mathrm{const}_j = \frac{1}{2} + \frac{j-2}{2} M_1^{(j)}$ for $j \geq 3$ (and the same formula at $j = 2$ would give $\frac{1}{2}$; the actual $\mathrm{const}_2 = \frac{3}{2} - 4c_2$ differs by the $a = 1$ off-diagonal contribution $1 - 4c_2$). Separating out $j = 2$:

$$d_n^* \;=\; \binom{n}{2}\mathrm{const}_2 + \sum_{j=3}^{n}(-1)^j \binom{n}{j}\left(\tfrac{1}{2} + \tfrac{j-2}{2} M_1^{(j)}\right).$$

which is precisely the symmetric form (26). For $n = 3$ this evaluates to $d_3^* = 3\,\mathrm{const}_2 - \mathrm{const}_3 = 3(\frac{3}{2} - 4c_2) - (\frac{2}{3} - c_3) = \frac{23}{6} - 12c_2 + c_3 \approx 1.273$. □

Table 10: Verification of Theorem 13 for $n = 1, 2, 3$ and $q = 10, 100, 1000, 10000$. "Exact" is computed via Theorem 6 $P_{\mathrm{nodes}}(n, q) = nq + (n-1) - \sum_{j=2}^{n}(-1)^j \binom{n}{j} S_j(q)$ with S_j from the brute-force (a, k) sum (Table 9). "Est" is $nq + (n-1) - c_n^*\sqrt{q} - d_n^*$. The residual Exact − Est decays at rate $\mathrm{O}(q^{-1/2})$ as expected. For $n = 1$ the columns c_n^* and d_n^* are 0 by convention (the empty sums $\sum_{j=2}^{1}$).

n	q	$P_{\mathrm{nodes}}(n, q)$ exact	est	residual	c_n^*	d_n^*
1	10	10.00000	10.00000	0.00000	0	0
1	100	100.00000	100.00000	0.00000	0	0
1	1000	1000.00000	1000.00000	0.00000	0	0
1	10000	10000.00000	10000.00000	0.00000	0	0
2	10	19.43909	19.68961	−0.25052	0.22634	0.59465
2	100	198.06841	198.14197	−0.07356	0.22634	0.59465
2	1000	1993.22512	1993.24790	−0.02278	0.22634	0.59465
2	10000	19977.76435	19977.77150	−0.00716	0.22634	0.59465
3	10	28.51051	29.07122	−0.56071	0.52371	1.27258
3	100	295.33062	295.49032	−0.15970	0.52371	1.27258
3	1000	2984.11716	2984.16639	−0.04923	0.52371	1.27258
3	10000	29948.34089	29948.35642	−0.01553	0.52371	1.27258

Coefficients used. $c_2 \approx 0.22634$, $c_3 \approx 0.15530$. Then

$$c_2^* = c_2, \qquad c_3^* = 3c_2 - c_3 \approx 0.52371,$$

$$d_2^* = \mathrm{const}_2 \approx 0.59465, \qquad d_3^* = 3\,\mathrm{const}_2 - \mathrm{const}_3 \;=\; \tfrac{23}{6} - 12c_2 + c_3 \;\approx\; 1.27258,$$

using $\mathrm{const}_3 = \frac{1}{2} + \frac{1}{2}(\frac{1}{3} - 2c_3) = \frac{2}{3} - c_3 \approx 0.51136$. The residuals shrink by $\sqrt{10}$ each time q scales by 10, confirming the $\mathrm{O}(q^{-1/2})$ rate of $R_n(q)$.

8 Combined: $P_{\mathbf{ptr}}(1, q)$

By Theorem 7,

$$P_{\mathrm{ptr}}(n, q) \;=\; 3P_{\mathrm{nodes}}(n, q) + 3 + W(n, q) + R(n, q),$$

where the general formulas are

$$W(n, q) \;=\; \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j} S_j^{(k)}(q), \qquad R(n, q) \;=\; \mathbf{E}\Big[\sum_{p \in \mathrm{Trie}} r_p\Big]$$

with $r_p \in \{-3, -1, 0\}$ the per-node correction from (1). For $n = 1$, each piece simplifies:

- $P_{\text{nodes}}(1,q) = q$ is exact (one key gives a path of q distinct prefixes; Theorem 13 at $n = 1$).
- $W(1,q) = \binom{1}{1} S_1^{(k)}(q) = S_1^{(k)}(q)$, which has the exact closed form (Lemma 5):

$$W(1,q) \;=\; S_1^{(k)}(q) \;=\; \frac{T_{q-2}}{T_q} \cdot \frac{q^3 - q}{6}.$$

- $R(1,q)$ has the exact closed form (Proposition 8):

$$R(1,q) \;=\; -\sum_{c=0}^{\lfloor (q-1)/2 \rfloor} \binom{q-1-c}{c} T_{q-1-2c} \frac{2(c+1)!}{T_q} \;-\; 3\sum_{c=0}^{\lfloor q/2 \rfloor} \binom{q-c}{c} T_{q-2c} \frac{c!}{T_q},$$

  the first sum from the $\text{TMS}_l = q-1-l$ boundary line ($r_p = -1$) and the second from $\text{TMS}_l = q-l$ ($r_p = -3$).

All the asymptotic content of $P_{\text{ptr}}(1,q)$ is therefore captured by the expansions of $W(1,q)$ (Lemma 14 below) and of $R(1,q)$ (Lemma 15).

Lemma 14 (Asymptotic of $W(1,q)$)**.** *As $q \to \infty$,*

$$W(1,q) \;=\; S_1^{(k)}(q) \;=\; \frac{q^2}{6} \;-\; \frac{q^{3/2}}{3} \;+\; \frac{q}{2} \;-\; \frac{7\sqrt{q}}{12} \;+\; \frac{5}{12} \;+\; \mathrm{O}(q^{-1/2}). \tag{27}$$

Proof. By Lemma 5, $W(1,q) = S_1^{(k)}(q) = \frac{q^3-q}{6} \cdot \frac{T_{q-2}}{T_q}$, so the proof reduces to expanding T_{q-2}/T_q to five orders. Setting $\rho(q) := T_{q-1}/T_q$, division of the recurrence $T_q = 2T_{q-1} + (q-1)T_{q-2}$ by T_q gives

$$\rho(q)[2 + (q-1)\rho(q-1)] = 1, \qquad \frac{T_{q-2}}{T_q} = \frac{1 - 2\rho(q)}{q-1}. \tag{28}$$

Wright's asymptotic guarantees $\rho(q)$ has a full asymptotic series $\rho(q) = \sum_{k\ge 1} a_k q^{-k/2}$, and (28) identifies the coefficients via dominant-balance coefficient matching. The computation (see §A.5) gives

$$\rho(q) = q^{-1/2} - q^{-1} + \tfrac{3}{4} q^{-3/2} - \tfrac{1}{4} q^{-2} - \tfrac{7}{32} q^{-5/2} + \mathrm{O}(q^{-3}),$$
$$\frac{T_{q-2}}{T_q} = \frac{1}{q} - \frac{2}{q^{3/2}} + \frac{3}{q^2} - \frac{7}{2q^{5/2}} + \frac{7}{2q^3} + \mathrm{O}(q^{-7/2}).$$

Multiplying T_{q-2}/T_q by $(q^3 - q)/6$: the q^0 term comes from $(q^3/6) \cdot (7/2) q^{-3} - (q/6) \cdot q^{-1} = 7/12 - 1/6 = 5/12$, yielding (27). □

Table 11: Verification of Lemma 14 for $q \in \{10, 100, 1000, 10000\}$. "Exact" is $W(1,q) = \frac{q^3-q}{6} \cdot T_{q-2}/T_q$ computed in exact rational arithmetic. "Est" is the four-term estimate $q^2/6 - q^{3/2}/3 + q/2 - 7\sqrt{q}/12$ (without the $+5/12$ constant). The residual Exact − Est converges to $5/12 \approx 0.4167$, the explicit $\mathrm{O}(1)$ constant of (27).

q	$W(1,q)$ exact	est (4-term)	residual
10	9.640603	9.281079	0.359524
100	1,377.898643	1,377.500000	0.398643
1000	156,607.705541	156,607.294513	0.411028
10000	16,338,275.414892	16,338,275.000000	0.414892

Lemma 15 (Asymptotic of $R(1,q)$)**.** *As $q \to \infty$,*

$$R(1,q) \;=\; (c_1 - 2)\sqrt{q} \;+\; (c_1 - 2) \;+\; \mathrm{O}(q^{-1/2}), \tag{29}$$

where $c_1 := \int_0^\infty e^{-\xi^2/2 - 2\xi}\, d\xi \approx 0.42136923$.

Proof. By Proposition 8, $R(1,q) = -B_1^{(1)}(q) - 3B_1^{(0)}(q)$ where

$$B_1^{(0)}(q) = \sum_{k=0}^{\lfloor q/2 \rfloor} \frac{h(q-k,k)}{T_q}, \qquad B_1^{(1)}(q) = 2 \sum_{k=0}^{\lfloor (q-1)/2 \rfloor} \frac{(k+1)\, h(q-1-k,k)}{T_q}.$$

Asymptotic of $B_1^{(0)}$. The $k=0$ term equals 1; for $k \geq 1$, apply Lemma 10 at $j=1, \alpha=0$ and Remark 11 (giving $j \int P\, e^{-\xi^2/2-2\xi} d\xi = -M_1^{(1)}/2 = -(1-2c_1)/2$ at $j=1$):

$$\sum_{k \geq 1} h(q-k,k)/T_q \;=\; c_1\sqrt{q} - \tfrac{1}{2} - \tfrac{1-2c_1}{2} + \mathrm{o}(1) \;=\; c_1\sqrt{q} + \mathrm{O}(1).$$

Tracking the constant explicitly, the $k=0$ contribution gives $+1$, the Euler–Maclaurin boundary half-correction at $k=0$ gives $-1/2$, and the profile sub-leading $P(\xi)/\sqrt{q}$ integrated against $e^{-\xi^2/2-2\xi}$ multiplied by $\sqrt{q}$ gives $\int_0^\infty P(\xi)\, e^{-\xi^2/2-2\xi} d\xi = -(1-2c_1)/2 \cdot 1 = c_1 - 1/2$ (using $M_1^{(1)} = 1-2c_1$ and the Remark 11 identity at $j=1$). Summing: $1 - 1/2 + (c_1 - 1/2) = c_1$, so

$$B_1^{(0)}(q) \;=\; c_1\sqrt{q} + c_1 + \mathrm{o}(1). \tag{30}$$

Asymptotic of $B_1^{(1)}$. Factor $h(q-1-k,k)/T_q = \rho(q) \cdot h(q-1-k,k)/T_{q-1}$, where the inner ratio is the size-$(q-1)$ profile (Lemma 9 at size $q-1$) and $\rho(q) = q^{-1/2} - q^{-1} + \mathrm{O}(q^{-3/2})$ (Lemma 14). At size $N := q-1$, apply Lemma 10 with the $\sqrt{N}$-coefficient supplied by Remark 11's identity $j \int \xi\, P(\xi)\, e^{-j\xi^2/2-2j\xi} d\xi = (j-3)\big[(4j+1)c_j - 2\big]/(2j)$ which at $j=1$ gives $2-5c_1$:

$$\sum_{k \geq 1} k \cdot \frac{h(N-k,k)}{T_N} = N M_1^{(1)} + \sqrt{N}(2-5c_1) + \mathrm{o}(\sqrt{N}) \;=\; N(1-2c_1) + \sqrt{N}(2-5c_1) + \mathrm{o}(\sqrt{N}),$$

$$\sum_{k \geq 0} \frac{h(N-k,k)}{T_N} = c_1\sqrt{N} + c_1 + \mathrm{o}(1) \quad \text{(as for } B_1^{(0)} \text{ at size } N\text{)}.$$

Adding:

$$\sum_{k \geq 0} (k+1)\, \frac{h(N-k,k)}{T_N} \;=\; N(1-2c_1) + \sqrt{N}(2-4c_1) + c_1 + \mathrm{o}(\sqrt{N}).$$

Multiplying by $2\rho(q) = 2q^{-1/2} - 2q^{-1} + \mathrm{O}(q^{-3/2})$, with $N = q-1$ (so $\sqrt{N} = \sqrt{q} - 1/(2\sqrt{q}) + \mathrm{O}(q^{-3/2})$):

$$\begin{aligned}
2\rho(q) \cdot N(1-2c_1) &= 2(1-2c_1)\big[(q-1)q^{-1/2} - (q-1)q^{-1} + \mathrm{O}(q^{-3/2}) \cdot q\big] \\
&= 2(1-2c_1)\big[\sqrt{q} - 1 - q^{-1/2} + \mathrm{O}(1) \cdot q^{-1}\big] \\
&= 2(1-2c_1)\sqrt{q} - 2(1-2c_1) + \mathrm{O}(q^{-1/2}), \\
2\rho(q) \cdot \sqrt{N}(2-4c_1) &= 2(2-4c_1)\big[1 + \mathrm{O}(q^{-1/2})\big] \;=\; (4-8c_1) + \mathrm{O}(q^{-1/2}).
\end{aligned}$$

The $2\rho(q) \cdot c_1$ contributes $\mathrm{O}(q^{-1/2})$. Summing:

$$B_1^{(1)}(q) \;=\; 2(1-2c_1)\sqrt{q} + (2-4c_1) + \mathrm{O}(q^{-1/2}), \tag{31}$$

where the constant collects $-2(1-2c_1) + (4-8c_1) = (2-4c_1)$.

Combine. $R(1,q) = -B_1^{(1)}(q) - 3B_1^{(0)}(q) = -[2(1-2c_1)\sqrt{q} + (2-4c_1)] - 3[c_1\sqrt{q} + c_1] + \mathrm{O}(q^{-1/2}) = (c_1-2)\sqrt{q} + (c_1-2) + \mathrm{O}(q^{-1/2})$ (both coefficients collapse to $c_1 - 2$). □

Table 12: Verification of Lemma 15 for $q \in \{10, 100, 1000, 10000\}$. "Exact" is $R(1,q) = -B_1^{(1)}(q) - 3B_1^{(0)}(q)$ from Proposition 8, evaluated in exact rational arithmetic by summing $h(q-l,k)/T_q$ term by term. "Est" is the leading estimate $(c_1-2)\sqrt{q}$ with $c_1 \approx 0.42136923$. The residual Exact − Est converges to a finite O(1) constant $c_1 - 2 \approx -1.579$, as predicted by (29).

q	$R(1,q)$ exact	est $((c_1-2)\sqrt{q})$	residual
10	−6.899707	−4.992069	−1.907638
100	−17.479101	−15.786308	−1.692793
1000	−51.536354	−49.920688	−1.615666
10000	−159.453508	−157.863077	−1.590431

Theorem 16 ($P_{\mathrm{ptr}}(1,q)$, four leading orders)**.** *As $q \to \infty$,*

$$\boxed{P_{ptr}(1,q) \;=\; \frac{q^2}{6} \;-\; \frac{q^{3/2}}{3} \;+\; \frac{7q}{2} \;+\; \left(c_1 - \tfrac{31}{12}\right)\sqrt{q} \;+\; \left(\tfrac{17}{12} + c_1\right) \;+\; \mathrm{O}(q^{-1/2}).} \tag{32}$$

Proof. By Theorem 7 and $P_{\mathrm{nodes}}(1,q) = q$, $P_{\mathrm{ptr}}(1,q) = 3q+3+W(1,q)+R(1,q)$. Substituting $W(1,q)$ from (27) (which carries through the constant $5/12$ from $W(1,q) = q^2/6 - q^{3/2}/3 + q/2 - 7\sqrt{q}/12 + 5/12 + \mathrm{O}(q^{-1/2})$) and $R(1,q)$ from (29),

$$\begin{aligned} P_{\mathrm{ptr}}(1,q) &= 3q+3+\left[\tfrac{q^2}{6} - \tfrac{q^{3/2}}{3} + \tfrac{q}{2} - \tfrac{7\sqrt{q}}{12} + \tfrac{5}{12}\right] + \left[(c_1-2)\sqrt{q} + (c_1-2)\right] + \mathrm{O}(q^{-1/2}) \\ &= \tfrac{q^2}{6} - \tfrac{q^{3/2}}{3} + \tfrac{7q}{2} + \left(c_1 - \tfrac{31}{12}\right)\sqrt{q} + \left(\tfrac{17}{12} + c_1\right) + \mathrm{O}(q^{-1/2}), \end{aligned}$$

where the q coefficient collects $3 + \frac{1}{2} = \frac{7}{2}$, the $\sqrt{q}$ coefficient combines $-\frac{7}{12} + (c_1 - 2) = c_1 - \frac{31}{12}$, and the constant collects $3 + \frac{5}{12} + (c_1 - 2) = \frac{17}{12} + c_1 \approx 1.83803$. □

Table 13: Verification of Theorem 16 for $q \in \{10, 100, 1000, 10000\}$. "Exact" is $P_{\mathrm{ptr}}(1,q) = 3q+3+W(1,q)+R(1,q)$ in exact rational arithmetic, with $W(1,q) = \frac{q^3-q}{6} \cdot T_{q-2}/T_q$ and $R(1,q) = -B_1^{(1)}(q) - 3B_1^{(0)}(q)$ summed term by term over k. "Est" is the four-term estimate $q^2/6 - q^{3/2}/3 + 7q/2 + (c_1 - \frac{31}{12})\sqrt{q}$ with $c_1 \approx 0.42136923$. The residual Exact − Est converges to $C^* = \frac{17}{12} + c_1 \approx 1.83803$, as predicted by (32).

q	$P_{\mathrm{ptr}}(1,q)$ exact	est (4-term)	residual
10	35.740896	34.289010	1.451886
100	1,663.419542	1,661.713692	1.705850
1000	159,559.169187	159,557.373825	1.795362
10000	16,368,118.961384	16,368,117.136923	1.824461

9 Combined: $P_{\mathrm{ptr}}(n,q)$ for general n

We close the general-n pointer count by carrying the boundary-layer machinery of §6 through to two new families of sums: the k-weighted moments $S_j^{(k)}(q)$ and the boundary moments $B_j^{(b)}(q)$ that make up $R(n,q)$. The decisive input is that the latter collapse to the $j=1$ term, so that $R(n,q) = n\,R(1,q)$ up to an exponentially small remainder of rate $-q\log 2$.

Boundary moments collapse for $j \geq 2$

Lemma 17 ($B_j^{(b)}$ exponential bound)**.** *For $b \in \{0,1\}$ and each fixed $j \geq 2$, the boundary moment*

$$B_j^{(b)}(q) \;:=\; \sum_{(l,k):\, q-l-k=b} f(l,k)\Big(\frac{h(q-l,k)}{T_q}\Big)^j \tag{33}$$

is exponentially small in q:

$$\log B_j^{(b)}(q) \;\leq\; -\tfrac{j\log 2}{2}\, q \;-\; 2j\sqrt{q} \;+\; \mathrm{O}(\log q), \tag{34}$$

equivalently $B_j^{(b)}(q) \leq \exp(-\frac{j\log 2}{2}\, q\,(1+\mathrm{O}((\log q)/q)))$. Consequently,

$$R(n,q) \;=\; n\cdot R(1,q) \;+\; \mathrm{O}\big(e^{-q\log 2\,(1+\mathrm{O}((\log q)/q))}\big). \tag{35}$$

Proof. Setup. On the boundary line $\{q-l-k=b\}$, $l=q-b-k$ and $h(q-l,k)=k!\binom{b+k}{k}T_b$ ($=k!$ for $b=0$, $=2(k+1)!$ for $b=1$). Substituting into (33):

$$B_j^{(b)}(q) = \sum_{k=0}^{\lfloor (q-b)/2\rfloor} \Phi_k, \qquad \Phi_k := \binom{q-b-k}{k} T_{q-b-2k}\cdot (h(b+k,k)/T_q)^j.$$

Quantitative bound (full details in §A.3; ratio-monotonicity approach). The consecutive ratio $\varrho_k := \Phi_{k+1}/\Phi_k$ (for $b=0$) simplifies via the recurrence $T_N = 2T_{N-1}+(N-1)T_{N-2}$ to

$$\frac{\Phi_{k+1}}{\Phi_k} = \frac{(q-2k)(k+1)^{j-1}(1-2\rho(q-2k))}{q-k}, \quad \rho(N) := T_{N-1}/T_N \leq 1/\sqrt{N}.$$

A three-regime case analysis (left endpoint $k=1$, right endpoint $k=\lfloor q/2\rfloor-2$, interior) shows $\varrho_k \geq 1$ for $1 \leq k \leq \lfloor q/2\rfloor - 2$ and $q \geq 20$, so $\Phi_1 \leq \Phi_2 \leq \cdots \leq \Phi_{\lfloor q/2\rfloor-1}$. At the edges the ratio can dip slightly below 1 ($\Phi_1/\Phi_0 = 1-2\rho(q) < 1$, and for $j=2$ also $\Phi_{q/2}/\Phi_{q/2-1} \to 2/5 < 1$), so the global maximum sits at $k^* \in \{\lfloor q/2\rfloor -1, \lfloor q/2\rfloor\}$ with $\Phi_{k^*} \leq \frac{5}{2}\Phi_{\lfloor q/2\rfloor}(1+\mathrm{o}(1))$ (worst-case constant from $j=2, q$ even). Direct Stirling+Wright evaluation of $\Phi_{\lfloor q/2\rfloor}$ (where $\binom{q/2}{q/2}T_0 = 1$, so $\Phi_{q/2} = ((q/2)!/T_q)^j$) gives $\log \Phi_{\lfloor q/2\rfloor} = -(jq\log 2)/2 - 2j\sqrt{q} + \mathrm{O}(\log q)$. Hence

$$B_j^{(0)}(q) \;\leq\; (q/2+1)\cdot \max_k \Phi_k \;\leq\; e^{-(j\log 2/2)q\cdot(1+\mathrm{o}(1))},$$

which is (34) for $b=0$. The case $b=1$ reduces to $b=0$ via the multiplicative identity $\Phi_k^{(1)}(q) = (2(k+1)\rho(q))^j \Phi_k^{(0)}(q-1)$, giving the same bound.

Inclusion–exclusion. By the derivation of $R(n,q)$ at (11) (just before Theorem 7):

$$R(n,q) \;=\; -\sum_{j=1}^{n} (-1)^{j+1}\binom{n}{j}\big[B_j^{(1)}(q) + 3B_j^{(0)}(q)\big].$$

The $j=1$ term gives $-n[B_1^{(1)} + 3B_1^{(0)}] = n\,R(1,q)$ by Proposition 8. The $j \geq 2$ terms each have $|B_j^{(b)}(q)| \leq e^{-(j\log 2/2)q(1+o(1))}$, the sum dominated by $j=2$:

$$\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}\big[B_j^{(1)} + 3B_j^{(0)}\big] \;=\; \mathrm{O}\big(2^n\cdot e^{-q\log 2(1+\mathrm{o}(1))}\big) \;=\; \mathrm{O}\big(e^{-q\log 2(1+\mathrm{o}(1))}\big),$$

yielding (35). □

Table 14: Verification of Lemma 17 for $j \in \{2,3\}$, $b \in \{0,1\}$, $q \in \{10, 100, 1000, 10000\}$. "$\log_{10} B_j^{(b)}$" is the actual value, computed in log-space float64 from (33). "Bound" is the $\log_{10}$ of the upper bound $\exp(-j(\log 2/2)q)$ in (34) (taking the $(1+\mathrm{o}(1))$ factor as 1, so the actual bound with sub-leading corrections is yet smaller). Both $b=0$ and $b=1$ columns show actual values well below the bound, with empirical decay rate slightly faster than $j \log 2/2$ due to the additional $-2j\sqrt{q}$ contribution from Wright's $e^{2\sqrt{q}}$ factor in T_q.

q	j	$\log_{10} B_j^{(0)}$	$\log_{10} B_j^{(1)}$	$\log_{10}$ bound
10	2	−4.63	−3.88	−3.01
10	3	−8.74	−7.03	−4.52
100	2	−43.05	−40.79	−30.10
100	3	−65.79	−62.90	−45.15
1000	2	−350.52	−347.24	−301.03
1000	3	−526.98	−523.08	−451.55
10000	2	−3177.57	−3173.28	−3010.30
10000	3	−4767.54	−4762.63	−4515.45

The k-weighted moments $S_j^{(k)}$ for $j \geq 2$

Let $M_n^{(j)} := \int_0^\infty \xi^n \exp\bigl(-(j/2)\xi^2 - 2j\xi\bigr) d\xi$ denote the n-th moment of $[h(q-k,k)/T_q]^j$'s leading profile $e^{-j\xi^2/2-2j\xi}$ (so $M_0^{(j)} = c_j$ and $M_1^{(j)} = 1/j - 2c_j$ via integration by parts of $d[e^{-j\xi^2/2-2j\xi}] = -(j\xi + 2j)e^{\cdots} d\xi$). The boundary-layer machinery of §6 now applies with the extra k-weight in the sum (giving an extra factor $k = \sqrt{q}\,\xi$ inside the Riemann integrand), yielding:

Lemma 18 ($S_j^{(k)}$ closed form)**.** *For each fixed integer $j \geq 2$, as $q \to \infty$,*

$$S_j^{(k)}(q) \;=\; c_j^{(k)}\, q \;+\; d_j^{(k)} \sqrt{q} \;+\; \mathrm{O}(1),$$

with

$$c_j^{(k)} = \frac{1}{j} - 2c_j \quad (j \geq 2),$$

$$d_j^{(k)} = \frac{(j-3)\bigl((4j+1)c_j - 2\bigr)}{2j} \quad (j \geq 4), \quad d_3^{(k)} = 0, \quad d_2^{(k)} = \tfrac{27}{4}c_2 - \tfrac{3}{2}.$$

Proof of Lemma 18. Setup. By definition,

$$S_j^{(k)}(q) \;=\; \sum_{l=1}^{q} \sum_{k=0}^{l} k\, f(l,k) \Bigl(\tfrac{h(q-l,k)}{T_q}\Bigr)^j.$$

Reparametrise via $a := l - k \geq 0$ (so $l = a + k$); the $k = 0$ rows contribute 0 (factor of k):

$$S_j^{(k)}(q) \;=\; \sum_{\substack{a,k \geq 0 \\ a+2k \leq q \\ k \geq 1}} k \binom{a+k}{k} T_a \Bigl[\frac{k!\binom{q-a-k}{k} T_{q-a-2k}}{T_q}\Bigr]^j.$$

Split by a: *diagonal* $a = 0$ and *off-diagonal slices* $a \geq 1$.

Diagonal $a = 0, k \geq 1$. On $a = 0$, $\binom{k}{k} T_0 = 1$ and the summand reduces to $k \cdot [h(q-k,k)/T_q]^j$. Apply Lemma 10 with $\alpha = 1$ and substitute the profile integral evaluation $j \int \xi P\, e^{-j\xi^2/2-2j\xi} d\xi = (j-3)[(4j+1)c_j - 2]/(2j)$ from Remark 11:

$$\Sigma_{a=0}^{(k)} \;:=\; \sum_{k \geq 1} k \Bigl[\frac{h(q-k,k)}{T_q}\Bigr]^j \;=\; \bigl(\tfrac{1}{j} - 2c_j\bigr)\, q \;+\; \frac{(j-3)\bigl[(4j+1)c_j - 2\bigr]}{2j} \sqrt{q} \;+\; \mathrm{o}(\sqrt{q}).$$

Off-diagonal $a = 1, k \geq 1$. On $a = 1$, $\binom{1+k}{k}T_1 = 2(k+1)$, and using the identity $h(q-1-k,k)/T_q = \varrho_{1,k} \cdot h(q-k,k)/T_q$ where

$$\varrho_{1,k} := \frac{\binom{q-1-k}{k}T_{q-1-2k}}{\binom{q-k}{k}T_{q-2k}} \;=\; \frac{q-2k}{q-k} \cdot \rho(q-2k) \;=\; \frac{1}{\sqrt{q}}\Big(1 - \frac{1}{\sqrt{q}} + \mathrm{O}\Big(\frac{\xi^2+1}{q}\Big)\Big)$$

on the boundary layer (cancellation of the linear-in-u piece of $(1-2u)^{1/2}/(1-u)$ leaves $1-u^2/2+\mathrm{O}(u^3)$ for the leading-ρ part, while the $-N^{-1}$ correction of $\rho(q-2k)$ contributes the $-1/\sqrt{q}$ shift; here $u = \xi/\sqrt{q}$, $\xi = k/\sqrt{q}$). Hence $\varrho_{1,k}^j = q^{-j/2}(1+\mathrm{O}(q^{-1/2}))$ uniformly on the boundary layer (since on $\xi \leq \sqrt{\log q}$, $(\xi^2+1)/q \leq (\log q + 1)/q \ll q^{-1/2}$), in particular $q^{-j/2}(1+\mathrm{o}(1))$.

$$\begin{aligned}
\Sigma_{a=1}^{(k)} &= 2\sum_{k\geq 1} k(k+1)\cdot \varrho_{1,k}^j \Big[\frac{h(q-k,k)}{T_q}\Big]^j \\
&= \frac{2}{q^{j/2}} \sum_{k\geq 1} k(k+1)\, e^{-j\xi^2/2-2j\xi}(1+\mathrm{o}(1)) \\
&= \frac{2}{q^{j/2}} \cdot \big[q^{3/2}\, M_2^{(j)} + \mathrm{O}(q)\big] \;=\; 2M_2^{(j)}\, q^{(3-j)/2} + \mathrm{O}(q^{1-j/2}),
\end{aligned}$$

where the Riemann-sum step uses $k(k+1) = k^2 + k$ and (with $\xi_k = k/\sqrt{q}$, $\Delta\xi = 1/\sqrt{q}$, Euler–Maclaurin error $\mathrm{O}(1)$):

$$\sum_{k\geq 1} k^2\, e^{-j\xi^2/2-2j\xi} = q\sum_{k\geq 1} \xi_k^2\, e^{\cdots} = q\cdot\big(\sqrt{q}\, M_2^{(j)} + \mathrm{O}(1)\big) = q^{3/2}\, M_2^{(j)} + \mathrm{O}(q),$$
$$\sum_{k\geq 1} k\, e^{-j\xi^2/2-2j\xi} = \sqrt{q}\sum_{k\geq 1} \xi_k\, e^{\cdots} = \sqrt{q}\cdot\big(\sqrt{q}\, M_1^{(j)} + \mathrm{O}(1)\big) = q\, M_1^{(j)} + \mathrm{O}(\sqrt{q}),$$

so $\sum_k k(k+1)e^{\cdots} = q^{3/2}M_2^{(j)} + qM_1^{(j)} + \mathrm{O}(\sqrt{q}) = q^{3/2}M_2^{(j)} + \mathrm{O}(q)$.

Off-diagonal $a \geq 2$. By the analogous computation, for fixed a: $\varrho_{a,k} = \prod_{i=0}^{a-1}(q-2k-i)\rho(q-2k-i)/(q-k-i) = \mathrm{O}(q^{-a/2})$, so $\varrho_{a,k}^j = \mathrm{O}(q^{-aj/2})$, and the prefactor $k\binom{a+k}{k}T_a = \mathrm{O}_a(k^{a+1})$ (with constant depending on a through $T_a/a!$). Riemann-summing gives each slice $\mathrm{O}_a(q^{(a+2-aj)/2})$. For $j \geq 2$ and $a \geq 2$: $(a+2-aj)/2 = (a(1-j)+2)/2 \leq (2-a)/2$, so the slice is $\mathrm{O}_a(q^{(2-a)/2})$, geometric decay in a:

$$\sum_{a\geq 2} \Sigma_a^{(k)} = \Sigma_{a=2}^{(k)}\big(1 + \mathrm{O}(q^{-1/2}) + \mathrm{O}(q^{-1}) + \cdots\big) = \Sigma_{a=2}^{(k)}(1+\mathrm{o}(1)) = \mathrm{O}(1).$$

The large-a tail (where $a + 2k$ approaches the boundary q) is dominated by the boundary moments $B_j^{(b)}(q)$ (Lemma 17), which are exponentially small in q; hence the tail contribution is negligible. Thus $\Sigma_{a\geq 2}^{(k)} = \mathrm{O}(1) = \mathrm{o}(\sqrt{q})$.

Combining. $S_j^{(k)}(q) = \Sigma_{a=0}^{(k)} + \Sigma_{a=1}^{(k)} + \Sigma_{a\geq 2}^{(k)}$.

Case $j \geq 4$**:** $\Sigma_{a=1}^{(k)} = 2M_2^{(j)}q^{(3-j)/2} = \mathrm{O}(q^{-1/2}) = \mathrm{o}(\sqrt{q})$, and $\Sigma_{a\geq 2}^{(k)} = \mathrm{o}(\sqrt{q})$. Hence $S_j^{(k)}(q) = \Sigma_{a=0}^{(k)} + \mathrm{o}(\sqrt{q})$, i.e.

$$c_j^{(k)} = \tfrac{1}{j} - 2c_j, \qquad d_j^{(k)} = \frac{(j-3)\big[(4j+1)c_j - 2\big]}{2j}.$$

Case $j = 3$**:** $\Sigma_{a=1}^{(k)} = 2M_2^{(3)}\, q^0 + \mathrm{O}(q^{-1/2}) = 2M_2^{(3)} + \mathrm{O}(q^{-1/2})$, which is $\mathrm{O}(1)$, not affecting the $\sqrt{q}$-coefficient. The diagonal $\sqrt{q}$ coefficient $(j-3)\big[(4j+1)c_j - 2\big]/(2j) = 0$ at $j = 3$. Hence $d_3^{(k)} = 0$. (A pleasant algebraic coincidence: the $(j-3)$ factor in the diagonal $\sqrt{q}$-coefficient vanishes at exactly the same j for which $\Sigma^{(k)}_{a=1}$ drops to O(1), so two independent mechanisms agree on $d^{(k)}_3 = 0$.)

Case $j = 2$: $\Sigma^{(k)}_{a=1} = 2M^{(2)}_2\sqrt{q} + \mathrm{O}(1)$, which contributes to the $\sqrt{q}$-coefficient. With $M^{(2)}_2 = 9c_2/2 - 1$:

$$\Sigma^{(k)}_{a=1}\big|_{\sqrt{q}\text{-part}} \;=\; 2(9c_2/2-1) \;=\; 9c_2-2.$$

The diagonal $\sqrt{q}$ coefficient at $j = 2$: $(j-3)\big[(4j+1)c_j-2\big]/(2j) = (-1)(9c_2-2)/4 = (2-9c_2)/4$. Adding:

$$d^{(k)}_2 \;=\; \frac{2-9c_2}{4} + (9c_2-2) \;=\; \frac{2-9c_2+36c_2-8}{4} \;=\; \frac{27c_2-6}{4} \;=\; \tfrac{27}{4}c_2 - \tfrac{3}{2}.$$

This completes the proof. □

Table 15: Verification of Lemma 18 for $j \in \{2, 3, 4, 5\}$ and $q \in \{10, 100, 1000, 10000\}$. "Exact" is $S^{(k)}_j(q)$ computed in log-space float64 from (6). "Est" is $c^{(k)}_j q + d^{(k)}_j\sqrt{q}$ with the closed-form coefficients from Lemma 18 (using $c_2 \approx 0.22634$, $c_3 \approx 0.15530$, $c_4 \approx 0.11833$, $c_5 \approx 0.09561$). The residual converges to a finite O(1) constant for each j.

j	q	$S^{(k)}_j(q)$ exact	est	residual
2	10	0.512268	0.561094	−0.048825
2	100	4.981249	5.010145	−0.028896
2	1000	48.178072	48.201590	−0.023519
2	10000	475.986128	476.008005	−0.021876
3	10	0.178807	0.227256	−0.048449
3	100	2.220460	2.272556	−0.052096
3	1000	22.672972	22.725557	−0.052585
3	10000	227.202953	227.255569	−0.052616
4	10	0.082772	0.138040	−0.055268
4	100	1.279381	1.349193	−0.069812
4	1000	13.321245	13.393254	−0.072009
4	10000	133.548204	133.620487	−0.072282
5	10	0.041524	0.092748	−0.051225
5	100	0.821375	0.893832	−0.072457
5	1000	8.756335	8.831894	−0.075559
5	10000	87.906545	87.982412	−0.075867

Assembled closed form for $P_{\mathbf{ptr}}(n, q)$

Theorem 19 ($P_{\mathrm{ptr}}(n, q)$ for general n)**.** *For each integer $n \geq 1$, as $q \to \infty$,*

$$\boxed{P_{ptr}(n,q) \;=\; \frac{n}{6}\,q^2 \;-\; \frac{n}{3}\,q^{3/2} \;+\; A_n\,q \;+\; B_n\sqrt{q} \;+\; \mathrm{O}(1),}$$

with closed-form coefficients

$$A_n = \frac{7n}{2} \;+\; \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}\Big(\frac{1}{j} - 2c_j\Big), \tag{36}$$

$$B_n = n\Big(c_1 - \frac{31}{12}\Big) \;-\; 3c^*_n \;+\; \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}\,d^{(k)}_j, \tag{37}$$

where $c_n^* = \sum_{j=2}^n (-1)^j \binom{n}{j} c_j$ *(see Theorem 13, eq. (25)). The two leading coefficients* $n/6$ *and* $-n/3$ *are linear in* n *and rational; the* c_j*-dependent content first appears at order* q *(through* c_j *in* A_n*) and then at order* $\sqrt{q}$ *(through* c_n^* *and* $d_j^{(k)}$*).*

Proof. By Theorem 7, $P_{\text{ptr}}(n,q) = 3P_{\text{nodes}}(n,q) + 3 + W(n,q) + R(n,q)$. We assemble the four-term expansion piece by piece.

Piece 1: $3P_{nodes}(n,q)+3$. By Theorem 13, $P_{\text{nodes}}(n,q) = nq+(n-1)-c_n^*\sqrt{q}-d_n^*+\mathrm{O}(q^{-1/2})$, so

$$3P_{\text{nodes}}(n,q) + 3 \;=\; 3nq + 3n - 3c_n^*\sqrt{q} - 3d_n^* + \mathrm{O}(q^{-1/2}) \;=\; 3nq + 3n - 3c_n^*\sqrt{q} + \mathrm{O}(1), \quad (38)$$

using $3(n-1)+3 = 3n$ and absorbing the explicit constant $-3d_n^*$ into $\mathrm{O}(1)$. This contributes 0 at q^2 and $q^{3/2}$, $3n$ at q, $-3c_n^*$ at $\sqrt{q}$.

Piece 2: $W(n,q)$. By the inclusion–exclusion formula for $W(n,q)$ derived in Theorem 7, $W(n,q) = \sum_{j=1}^n (-1)^{j+1}\binom{n}{j} S_j^{(k)}(q)$. Split into $j=1$ and $j \geq 2$ contributions.

$j=1$ *piece*: $n \cdot S_1^{(k)}(q)$. By Lemma 5, $S_1^{(k)}(q) = (q^3-q)/6 \cdot T_{q-2}/T_q$; and by Lemma 14's expansion $W(1,q) = S_1^{(k)}(q) = q^2/6 - q^{3/2}/3 + q/2 - 7\sqrt{q}/12 + \mathrm{O}(1)$. Hence

$$n\,S_1^{(k)}(q) \;=\; \tfrac{n}{6}q^2 - \tfrac{n}{3}q^{3/2} + \tfrac{n}{2}q - \tfrac{7n}{12}\sqrt{q} + \mathrm{O}(1). \quad (39)$$

$j \geq 2$ *pieces*: by Lemma 18, $S_j^{(k)}(q) = (\tfrac{1}{j} - 2c_j)q + d_j^{(k)}\sqrt{q} + \mathrm{O}(1)$. So

$$\sum_{j=2}^n (-1)^{j+1}\binom{n}{j} S_j^{(k)}(q) \;=\; \Big[\sum_{j=2}^n (-1)^{j+1}\binom{n}{j}\big(\tfrac{1}{j} - 2c_j\big)\Big]q \;+\; \Big[\sum_{j=2}^n (-1)^{j+1}\binom{n}{j} d_j^{(k)}\Big]\sqrt{q} \;+\; \mathrm{O}(1). \quad (40)$$

Adding (39) and (40):

$$W(n,q) \;=\; \tfrac{n}{6}q^2 - \tfrac{n}{3}q^{3/2} + \Big[\tfrac{n}{2} + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j}\big(\tfrac{1}{j} - 2c_j\big)\Big]q + \Big[-\tfrac{7n}{12} + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j} d_j^{(k)}\Big]\sqrt{q} + \mathrm{O}(1). \quad (41)$$

Piece 3: $R(n,q)$. By the collapse identity (35) (Lemma 17), and Lemma 15 giving $R(1,q) = (c_1-2)\sqrt{q} + \mathrm{O}(1)$. Hence

$$R(n,q) \;=\; n(c_1-2)\sqrt{q} + \mathrm{O}(1). \quad (42)$$

Combine. Summing (38), (41), (42):

$$P_{\text{ptr}}(n,q) \;=\; \tfrac{n}{6}q^2 - \tfrac{n}{3}q^{3/2} + A_n\,q + B_n\sqrt{q} + \mathrm{O}(1),$$

where collecting the q and $\sqrt{q}$ coefficients:

$$A_n = 3n + \tfrac{n}{2} + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j}\big(\tfrac{1}{j} - 2c_j\big) = \tfrac{7n}{2} + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j}\big(\tfrac{1}{j} - 2c_j\big),$$

$$B_n = -3c_n^* - \tfrac{7n}{12} + n(c_1-2) + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j} d_j^{(k)} = n\big(c_1 - \tfrac{31}{12}\big) - 3c_n^* + \sum_{j=2}^n (-1)^{j+1}\binom{n}{j} d_j^{(k)},$$

using $c_1 - 2 - 7/12 = c_1 - 31/12$. These match (36), (37). □

Theorem 20 (Simplified form of A_n)**.** *The* q*-coefficient* A_n *of Theorem 19 admits the remarkably clean closed form*

$$\boxed{A_n \;=\; \tfrac{5n}{2} \;+\; H_n \;+\; 2c_n^*,} \quad (43)$$

where $H_n := \sum_{k=1}^{n} 1/k$ is the n-th harmonic number and $c_n^* := \sum_{j=2}^{n}(-1)^j\binom{n}{j}c_j$ is the same constant appearing in Theorem 13. The appearance of H_n reveals an entropy-like structure: the q-order term of $P_{ptr}(n,q)$ decomposes as a rational $\frac{5n}{2}+H_n$ (uniform expansion fed by the keystone identity) plus twice the boundary-layer prefix-coincidence cost c_n^*.

Proof. The classical identity $\sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}/j = H_n$ (equivalently, $\int_0^1(1-(1-x)^n)/x\,dx = H_n$ by expanding $(1-x)^n$ binomially) splits off the $j=1$ term to give $\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}/j = H_n - n$. Substituting into (36):

$$A_n \;=\; \tfrac{7n}{2} + (H_n - n) - 2\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}c_j \;=\; \tfrac{5n}{2} + H_n + 2c_n^*,$$

since $\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}c_j = -c_n^*$. ☐

Corollary 21 (Closed form for B_n)**.** *The $\sqrt{q}$-coefficient B_n in Theorem 19 admits the closed form*

$$\boxed{\begin{aligned} B_n &= nc_1 + 3H_n + n^2 - \tfrac{67n}{12} - 1 \\ &\quad - \tfrac{39}{4}\binom{n}{2}c_2 \;+\; 3\binom{n}{3}c_3 \;+\; \sum_{j=4}^{n}(-1)^{j+1}\binom{n}{j}\frac{4j^2-5j-3}{2j}\,c_j. \end{aligned}} \tag{44}$$

Equivalently, $B_n = nc_1 + B_n^{\mathrm{rat}} + \sum_{j=2}^{n}\gamma_j(n)\,c_j$, with $B_n^{\mathrm{rat}} := 3H_n + n^2 - 67n/12 - 1$, and weights

$$\gamma_2(n) = -\tfrac{39}{4}\binom{n}{2}, \quad \gamma_3(n) = 3\binom{n}{3}, \quad \gamma_j(n) = (-1)^{j+1}\binom{n}{j}\tfrac{4j^2-5j-3}{2j} \quad (j \ge 4).$$

Unlike A_n (Theorem 20), the c_j-dependent sum does not telescope to a single transcendental combination — the weights $(4j^2-5j-3)/(2j) = 2j - 5/2 - 3/(2j)$ mix j, 1, $1/j$ contributions, and we are unaware of a further reduction. The $j=2$ coefficient $\gamma_2(n)$ acquires an extra $-9\binom{n}{2}c_2$ relative to the general formula (giving the apparent weight $-39/4$ in place of $(2\cdot 2 - 5/2 - 3/4) = 3/4$): the source is the off-diagonal $a=1$ slice contribution $2M_2^{(2)} = 9c_2 - 2$ to $d_2^{(k)}$ (cf. Lemma 18, Case $j=2$), which is multiplied by $(-1)^{j+1}\binom{n}{j} = -\binom{n}{2}$ in the sum defining B_n.

Proof. From Theorem 19, $B_n = n(c_1 - 31/12) - 3c_n^* + \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}d_j^{(k)}$.

Rational part. The rational part of $d_j^{(k)}$ (from Lemma 18) is $-3/2$ at $j=2$, 0 at $j=3$, and $-(j-3)/j = -1 + 3/j$ for $j \ge 4$. Writing $\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j} = 1 - n$ (from $\sum_{j=0}^{n}(-1)^j\binom{n}{j} = 0$, subtracting $j = 0, 1$) and using the classical identity $\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}/j = H_n - n$ (used in Theorem 20):

$$\sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}\big[\mathrm{rat}(d_j^{(k)})\big] \;=\; 3H_n + n^2 - 3n - 1,$$

which combines with $-31n/12$ from $n(c_1 - 31/12)$ to give $B_n^{\mathrm{rat}} = 3H_n + n^2 - 67n/12 - 1$.

c_j coefficients. For each $j \ge 2$, sum the c_j contributions from $-3c_n^*$ and from $(-1)^{j+1}\binom{n}{j}d_j^{(k)}$:

- $j=2$: $-3\binom{n}{2} + (-\binom{n}{2})\cdot\frac{27}{4} = -(3 + \frac{27}{4})\binom{n}{2} = -\frac{39}{4}\binom{n}{2}$.
- $j=3$: $3\binom{n}{3} + 0 = 3\binom{n}{3}$ (using $d_3^{(k)} = 0$).
- $j \ge 4$: $3(-1)^{j+1}\binom{n}{j} + (-1)^{j+1}\binom{n}{j}\cdot\frac{(j-3)(4j+1)}{2j} = (-1)^{j+1}\binom{n}{j}\cdot\frac{6j+(j-3)(4j+1)}{2j} = (-1)^{j+1}\binom{n}{j}\cdot\frac{4j^2-5j-3}{2j}$.

The c_1 contribution is nc_1 from $n(c_1 - 31/12)$ alone (neither c_n^* nor any $d_j^{(k)}$ involves c_1). Assembling all parts yields (44). ☐

Example 22 (Worked $n = 1, 2, 3$)**.** Using $c_1 \approx 0.42136923$, $c_2 \approx 0.22634$, $c_3 \approx 0.15530$, $d_2^{(k)} = \frac{27}{4}c_2 - \frac{3}{2} \approx 0.02780$, $d_3^{(k)} = 0$:

- $n = 1$: $c_1^* = 0$ (the $\sum_{j=2}^{n}$ sums are empty when $n = 1$), $A_1 = \frac{7}{2} = 3.5$, $B_1 = c_1 - \frac{31}{12} \approx -2.1620$.
  (Recovers Theorem 16.)
- $n = 2$: $c_2^* = c_2$, $\binom{2}{2}\big(\frac{1}{2} - 2c_2\big) = \frac{1}{2} - 2c_2$
  $A_2 = 7 - \big(\frac{1}{2} - 2c_2\big) = \frac{13}{2} + 2c_2 \approx 6.9527$, $B_2 = 2\big(c_1 - \frac{31}{12}\big) - 3c_2 - d_2^{(k)} \approx -5.0308$.
- $n = 3$: $c_3^* = 3c_2 - c_3 \approx 0.5237$, $d_3^{(k)} = 0$
  $A_3 = \frac{21}{2} - 3\big(\frac{1}{2} - 2c_2\big) + \big(\frac{1}{3} - 2c_3\big) \approx 10.3808$, $B_3 = 3\big(c_1 - \frac{31}{12}\big) - 3(3c_2 - c_3) - 3\,d_2^{(k)} \approx -8.1405$.

Substituting into the boxed display of Theorem 19, e.g. for $n = 3$:

$$P_{\text{ptr}}(3, q) \;\approx\; \tfrac{1}{2}q^2 - q^{3/2} + 10.3808\,q - 8.1405\sqrt{q} + \mathrm{O}(1),$$

matching Table 16 at every q to O(1).

Table 16: Verification of Theorem 19 for $n \in \{1, 2, 3\}$ and $q \in \{10, 100, 1000, 10000\}$. "Numeric" is $P_{\text{ptr}}(n, q) = 3P_{\text{nodes}}(n, q) + 3 + W(n, q) + R(n, q)$ computed in log-space `float64` from (9); at $q = 10000$ this loses about 10^{-3} of absolute precision relative to the exact rational value (compare Table 13's 16,368,118.961384 versus the entry 16,368,118.962003 below). "Est" is $\frac{n}{6}q^2 - \frac{n}{3}q^{3/2} + A_n q + B_n\sqrt{q}$ with A_n, B_n from (36)–(37). The residual converges to a finite O(1) constant (≈ 1.8245 for $n = 1$, ≈ 1.83 for $n = 2$, ≈ 1.62 for $n = 3$ as q grows).

n	q	$P_{\text{ptr}}(n, q)$ numeric	est	residual
1	10	35.740896	34.289010	1.4519
1	100	1,663.419542	1,661.713692	1.7058
1	1000	159,559.169187	159,557.373825	1.7954
1	10000	16,368,118.962003	16,368,117.136923	1.8251
2	10	66.286984	65.869691	0.4173
2	100	3,313.063072	3,311.627083	1.4360
2	1000	319,046.835662	319,045.073702	1.7620
2	10000	32,735,692.230917	32,735,690.364284	1.8666
3	10	95.396820	96.442641	−1.0458
3	100	4,957.527195	4,956.671846	0.8553
3	1000	478,501.997783	478,500.558608	1.4392
3	10000	49,102,995.153242	49,102,993.528818	1.6244

10 Summary of Part II

Part II established four-term asymptotic expansions for both $P_{\text{nodes}}(n, q)$ and $P_{\text{ptr}}(n, q)$, with all coefficients in closed form (rational + $\mathbb{Q}$-linear combinations of the $c_j := \int_0^\infty e^{-(j/2)\xi^2 - 2j\xi} d\xi$):

$$\begin{aligned} P_{\text{nodes}}(n, q) &= nq + (n-1) - c_n^*\sqrt{q} - d_n^* + \mathrm{O}(q^{-1/2}), && \text{Theorem 13,} \\ P_{\text{ptr}}(1, q) &= \tfrac{q^2}{6} - \tfrac{q^{3/2}}{3} + \tfrac{7q}{2} + \big(c_1 - \tfrac{31}{12}\big)\sqrt{q} + \mathrm{O}(1), && \text{Theorem 16,} \\ P_{\text{ptr}}(n, q) &= \tfrac{n}{6}q^2 - \tfrac{n}{3}q^{3/2} + A_n q + B_n\sqrt{q} + \mathrm{O}(1), && \text{Theorem 19.} \end{aligned}$$

The coefficients are

$$c_n^* = \sum_{j=2}^{n}(-1)^j\binom{n}{j}c_j, \quad A_n = \tfrac{7n}{2} + \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j}\big(\tfrac{1}{j} - 2c_j\big),$$

$$d_n^* = \binom{n}{2}(\mathrm{const}_2 - \tfrac{1}{2}) + \tfrac{n-1}{2} + \sum_{j=3}^{n}(-1)^j \binom{n}{j} \tfrac{j-2}{2} M_1^{(j)}, \quad B_n = n\bigl(c_1 - \tfrac{31}{12}\bigr) - 3c_n^* + \sum_{j=2}^{n}(-1)^{j+1}\binom{n}{j} d_j^{(k)};$$

the latter admits an explicit closed form $B_n = nc_1 + 3H_n + n^2 - 67n/12 - 1 + \sum_{j=2}^{n} \gamma_j(n)c_j$ with rational weights $\gamma_j(n)$ (Corollary 21).

The proofs reduced to three asymptotic facts about the building blocks of Part I:

(A) *Leading constants of S_j* (Lemma 12): $S_j(q) = c_j\sqrt{q} + \mathrm{const}_j + \mathrm{O}(q^{-1/2})$, with $\mathrm{const}_2 = \frac{3}{2} - 4c_2$ and $\mathrm{const}_j = \frac{1}{2} + \frac{j-2}{2}(\frac{1}{j} - 2c_j)$ for $j \geq 3$.

(B) *Asymptotic of $\rho(q) = T_{q-1}/T_q$* (Lemma 14): bootstrapped from the recurrence $\rho(q)\bigl[2 + (q-1)\rho(q-1)\bigr] = 1$ to $\rho(q) = q^{-1/2} - q^{-1} + \frac{3}{4}q^{-3/2} + \mathrm{O}(q^{-2})$, yielding T_{q-2}/T_q to four orders and hence $W(1,q)$ via the closed form $S_1^{(k)}(q) = (q^3 - q)T_{q-2}/(6T_q)$ (Lemma 5).

(C) *Boundary-layer machinery*: the profile lemma (Lemma 9) gives

$$h(q-k,k)/T_q = e^{-\xi^2/2 - 2\xi}\bigl(1 + \mathrm{O}((\xi^3 + 1)/\sqrt{q})\bigr)$$

uniformly on the boundary layer $k \leq \sqrt{q \log q}$; this is the single analytical workhorse driving (A), the $R(1,q)$ asymptotic (Lemma 15), and the $S_j^{(k)}$ closed form (Lemma 18). The boundary moments $B_j^{(b)}(q)$ for $j \geq 2$ are exponentially small (Lemma 17), so (35) reduces the general-n pointer count to the $n = 1$ case.

The Wright saddle-point asymptotic for T_N (eq. (15)) is the sole external analytical input. Wright's constant C cancels in every ratio $h(q-l,k)/T_q$ formed in this analysis, so the closed-form c_j, $c_j^{(k)}$, $d_j^{(k)}$ are independent of its value.

Each result was verified numerically in its own section (Tables 8, 10, 11, 12, 13, 14, 15, 16).

A Stirling–Wright Bookkeeping

This appendix collects every Stirling-and-Wright manipulation extracted from the main paper, written out at full detail: each cancellation is performed term by term, each Taylor expansion to as many orders as needed, each $\log q$ coefficient verified by explicit arithmetic. The main paper states its Lemmas and Theorems and uses their consequences; readers wanting the bookkeeping come here.

Conventions. q is the key length, k the open-pair count of a prefix, with $\xi = k/\sqrt{q}$ and $u = k/q$. The marked-involution count $T_N = |\{\text{length-}N \text{ marked involutions}\}|$ (OEIS A005425) satisfies the EGF $\sum_{N\geq 0} T_N z^N/N! = e^{2z+z^2/2}$ and the recurrence $T_{N+2} = 2T_{N+1} + (N+1)T_N$. We use Stirling's formula

$$\log m! = (m + \tfrac{1}{2})\log m - m + \tfrac{1}{2}\log(2\pi) + \mathrm{O}(1/m) \tag{45}$$

and Wright's saddle-point asymptotic

$$\log T_N = \tfrac{N}{2}\log N - \tfrac{N}{2} + 2\sqrt{N} - \tfrac{1}{4} + \log C + r(N), \tag{46}$$

where C is an absolute positive constant and the remainder $r(N)$ satisfies $r(N) \to 0$ as $N \to \infty$, with the uniform bound $|r(N)| \leq K_0$ for all $N \geq 1$ ($K_0 \leq 1$ suffices; numerical value ≈ 0.54).

Quantity expanded. The recurring object is $h(q-k,k)/T_q = k!\binom{q-k}{k}T_{q-2k}/T_q$, the probability (under Hayman's measure) that the first k canonical-prefix positions of a random marked involution are all "open new pair".

A.1 Boundary-layer expansion of $h(q-k,k)/T_q$

A.1.1 Stirling+Wright decomposition

By the closed form $h(q-k,k) = k!\binom{q-k}{k}T_{q-2k}$:

$$\log \frac{h(q-k,k)}{T_q} \;=\; \log k! + \log \binom{q-k}{k} + \log T_{q-2k} - \log T_q.$$

Expanding $\binom{q-k}{k} = (q-k)!/(k!(q-2k)!)$ cancels the $\log k!$:

$$\log \frac{h(q-k,k)}{T_q} \;=\; \log \frac{(q-k)!}{(q-2k)!} \;+\; \log T_{q-2k} \;-\; \log T_q. \tag{47}$$

On the boundary layer $1 \le k \le \sqrt{q\log q}$, $q-2k \ge q - 2\sqrt{q\log q} \ge q/2$ for q large enough that $4\sqrt{\log q} \le \sqrt{q}$ (i.e. $q \ge$ absolute constant), so Stirling applied to $(q-k)!$ and $(q-2k)!$ has $\mathrm{O}(1/q)$ remainder, and Wright applied to T_{q-2k}, T_q has $\mathrm{O}(q^{-1/2})$ remainder. Apply Stirling to each factorial:

$$\begin{aligned}\log(q-k)! &= (q-k+\tfrac12)\log(q-k) - (q-k) + \tfrac12\log(2\pi) + \mathrm{O}(q^{-1}),\\ \log(q-2k)! &= (q-2k+\tfrac12)\log(q-2k) - (q-2k) + \tfrac12\log(2\pi) + \mathrm{O}(q^{-1}).\end{aligned}$$

Subtracting:

$$\begin{aligned}\log\frac{(q-k)!}{(q-2k)!} &= (q-k+\tfrac12)\log(q-k) - (q-2k+\tfrac12)\log(q-2k)\\ &\quad - \big[(q-k)-(q-2k)\big] + \big[\tfrac12\log(2\pi) - \tfrac12\log(2\pi)\big] + \mathrm{O}(q^{-1})\\ &= (q-k+\tfrac12)\log(q-k) - (q-2k+\tfrac12)\log(q-2k) - k + \mathrm{O}(q^{-1}),\end{aligned} \tag{48}$$

the two $\frac12\log(2\pi)$ pieces cancelling. Apply Wright to T_{q-2k}, T_q:

$$\begin{aligned}\log T_{q-2k} &= \tfrac{q-2k}{2}\log(q-2k) - \tfrac{q-2k}{2} + 2\sqrt{q-2k} - \tfrac14 + \log C + r(q-2k),\\ \log T_q &= \tfrac{q}{2}\log q - \tfrac{q}{2} + 2\sqrt{q} - \tfrac14 + \log C + r(q).\end{aligned}$$

Subtracting:

$$\begin{aligned}\log T_{q-2k} - \log T_q &= \tfrac{q-2k}{2}\log(q-2k) - \tfrac{q}{2}\log q\\ &\quad - \big[\tfrac{q-2k}{2} - \tfrac{q}{2}\big] + \big[2\sqrt{q-2k} - 2\sqrt{q}\big]\\ &\quad + \big[-\tfrac14 + \tfrac14\big] + \big[\log C - \log C\big] + \big[r(q-2k) - r(q)\big]\\ &= \tfrac{q-2k}{2}\log(q-2k) - \tfrac{q}{2}\log q + k + 2(\sqrt{q-2k} - \sqrt{q}) + \mathrm{O}(q^{-1/2}),\end{aligned} \tag{49}$$

the $-\frac14$ pieces and the $\log C$ pieces cancelling, and the remainders bounded by $|r(q-2k)| + |r(q)| \le 2K_0$ (which is $\mathrm{O}(q^{-1/2})$ since for $q-2k \ge q/2$ we have $r(q-2k) = \mathrm{O}(q^{-1/2})$ via Wright's expansion).

Now add (48) and (49) via (47). The $-k$ from (48) cancels the $+k$ from (49):

$$\big[-k\big] + \big[+k\big] = 0.$$

The two $\log(q-2k)$-terms combine:

$$\underbrace{-(q-2k+\tfrac12)\log(q-2k)}_{\text{from (48)}} + \underbrace{\tfrac{q-2k}{2}\log(q-2k)}_{\text{from (49)}} = \big[-(q-2k+\tfrac12) + \tfrac{q-2k}{2}\big]\log(q-2k).$$

The bracket simplifies: $-(q-2k+\frac{1}{2})+\frac{q-2k}{2} = -(q-2k)-\frac{1}{2}+\frac{q-2k}{2} = -\frac{q-2k}{2}-\frac{1}{2}$. Hence

$$-(q-2k+\tfrac{1}{2})\log(q-2k)+\tfrac{q-2k}{2}\log(q-2k) \;=\; -\tfrac{q-2k}{2}\log(q-2k)-\tfrac{1}{2}\log(q-2k).$$

Putting it all together:

$$\begin{aligned}\log\frac{h(q-k,k)}{T_q} &= (q-k+\tfrac{1}{2})\log(q-k)-\tfrac{q-2k}{2}\log(q-2k)\\ &\quad -\tfrac{1}{2}\log(q-2k)-\tfrac{q}{2}\log q+2(\sqrt{q-2k}-\sqrt{q})+\mathrm{O}(q^{-1/2}).\end{aligned} \tag{50}$$

A.1.2 Taylor expansion of the components

Let $\xi = k/\sqrt{q}$, so $k=\xi\sqrt{q}$ and $u=k/q=\xi/\sqrt{q}$. We need three Taylor series; the standard ones are

$$\begin{aligned}\log(1-v) &= -v-\tfrac{v^2}{2}-\tfrac{v^3}{3}-\tfrac{v^4}{4}-\mathrm{O}(v^5),\\ \sqrt{1-v} &= 1-\tfrac{v}{2}-\tfrac{v^2}{8}-\tfrac{v^3}{16}-\tfrac{5v^4}{128}-\mathrm{O}(v^5).\end{aligned}$$

Applied with $v=u=\xi/\sqrt{q}$ to $\log(q-k)=\log q+\log(1-u)$:

$$\log(q-k) \;=\; \log q \;-\; \tfrac{\xi}{\sqrt{q}} \;-\; \tfrac{\xi^2}{2q} \;-\; \tfrac{\xi^3}{3q^{3/2}} \;-\; \tfrac{\xi^4}{4q^2} \;-\; \mathrm{O}(\xi^5/q^{5/2}). \tag{51}$$

With $v=2u=2\xi/\sqrt{q}$ to $\log(q-2k)=\log q+\log(1-2u)$:

$$\log(q-2k) \;=\; \log q \;-\; \tfrac{2\xi}{\sqrt{q}} \;-\; \tfrac{2\xi^2}{q} \;-\; \tfrac{8\xi^3}{3q^{3/2}} \;-\; \tfrac{4\xi^4}{q^2} \;-\; \mathrm{O}(\xi^5/q^{5/2}). \tag{52}$$

With $v=2u$ to $\sqrt{q-2k}=\sqrt{q}\,\sqrt{1-2u}$:

$$\sqrt{q-2k} \;=\; \sqrt{q} \;-\; \xi \;-\; \tfrac{\xi^2}{2\sqrt{q}} \;-\; \tfrac{\xi^3}{2q} \;-\; \tfrac{5\xi^4}{8q^{3/2}} \;-\; \mathrm{O}(\xi^5/q^2). \tag{53}$$

On the boundary layer ($\xi \le \sqrt{\log q}$), each Taylor remainder is dominated by any negative power of q (i.e. $\mathrm{o}(q^{-a})$ for every $a>0$), so all are $\mathrm{o}(1/\sqrt{q})$.

A.1.3 Multiplying by prefactors (Pieces 1–5)

We now substitute (51), (52), (53) into (50), mark the five summands as Pieces 1–5, and expand each.

Piece 1: $(q-k+\frac{1}{2})\log(q-k)=(q-\xi\sqrt{q}+\frac{1}{2})\log(q-k)$. Distribute across the three subfactors q, $-\xi\sqrt{q}$, $\frac{1}{2}$:

$$\begin{aligned}q\cdot(51) &= q\log q-\tfrac{q\xi}{\sqrt{q}}-\tfrac{q\xi^2}{2q}-\tfrac{q\xi^3}{3q^{3/2}}-\tfrac{q\xi^4}{4q^2}+\mathrm{O}(\xi^5/q^{3/2})\\ &= q\log q-\xi\sqrt{q}-\tfrac{\xi^2}{2}-\tfrac{\xi^3}{3\sqrt{q}}-\tfrac{\xi^4}{4q}+\mathrm{O}(\xi^5/q^{3/2}),\\ -\xi\sqrt{q}\cdot(51) &= -\xi\sqrt{q}\log q+\tfrac{\xi^2\sqrt{q}}{\sqrt{q}}+\tfrac{\xi^3\sqrt{q}}{2q}+\tfrac{\xi^4\sqrt{q}}{3q^{3/2}}+\tfrac{\xi^5\sqrt{q}}{4q^2}+\mathrm{O}(\xi^6/q^2)\\ &= -\xi\sqrt{q}\log q+\xi^2+\tfrac{\xi^3}{2\sqrt{q}}+\tfrac{\xi^4}{3q}+\mathrm{O}(\xi^5/q^{3/2}),\\ \tfrac{1}{2}\cdot(51) &= \tfrac{1}{2}\log q-\tfrac{\xi}{2\sqrt{q}}-\tfrac{\xi^2}{4q}+\mathrm{O}(\xi^3/q^{3/2}).\end{aligned}$$

Summing the three rows term-by-term:

$$\begin{aligned}(q-k+\tfrac{1}{2})\log(q-k) &= \underbrace{q\log q}_{(\dagger)} \;-\; \underbrace{\xi\sqrt{q}\log q}_{(\ddagger)} \;-\; \underbrace{\xi\sqrt{q}}_{(\S)} \;+\; \underbrace{\tfrac{1}{2}\log q}_{(*)}\\ &\qquad +\; \xi^2\big(-\tfrac{1}{2}+1\big) \;+\; \tfrac{\xi^3}{\sqrt{q}}\big(-\tfrac{1}{3}+\tfrac{1}{2}\big) \;-\; \tfrac{\xi}{2\sqrt{q}}\\ &\qquad +\; \mathrm{O}\big(\tfrac{\xi^4+1}{q}\big)\\ &= (\dagger)-(\ddagger)-(\S)+(*)+\tfrac{\xi^2}{2}+\tfrac{\xi^3}{6\sqrt{q}}-\tfrac{\xi}{2\sqrt{q}}+\mathrm{O}\big(\tfrac{\xi^4+1}{q}\big).\end{aligned} \tag{54}$$

Piece 2: $-\frac{q-2k}{2}\log(q-2k) = (-\frac{q}{2}+\xi\sqrt{q})\log(q-2k)$.

$$\begin{aligned}
-\tfrac{q}{2}\cdot(52) &= -\tfrac{q}{2}\log q + \tfrac{q\xi}{\sqrt{q}} + \tfrac{q\xi^2}{q} + \tfrac{4q\xi^3}{3q^{3/2}} + \tfrac{2q\xi^4}{q^2} + \mathrm{O}(\xi^5/q^{3/2}) \\
&= -\tfrac{q}{2}\log q + \xi\sqrt{q} + \xi^2 + \tfrac{4\xi^3}{3\sqrt{q}} + \tfrac{2\xi^4}{q} + \mathrm{O}(\xi^5/q^{3/2}), \\
\xi\sqrt{q}\cdot(52) &= \xi\sqrt{q}\log q - \tfrac{2\xi^2\sqrt{q}}{\sqrt{q}} - \tfrac{2\xi^3\sqrt{q}}{q} - \tfrac{8\xi^4\sqrt{q}}{3q^{3/2}} + \mathrm{O}(\xi^5/q^2) \\
&= \xi\sqrt{q}\log q - 2\xi^2 - \tfrac{2\xi^3}{\sqrt{q}} - \tfrac{8\xi^4}{3q} + \mathrm{O}(\xi^5/q^{3/2}).
\end{aligned}$$

Summing:

$$\begin{aligned}
-\tfrac{q-2k}{2}\log(q-2k) = {} & \underbrace{-\tfrac{q}{2}\log q}_{(\dagger)} + \underbrace{\xi\sqrt{q}\log q}_{(\ddagger)} + \underbrace{\xi\sqrt{q}}_{(\S)} \\
& + \xi^2(1-2) + \tfrac{\xi^3}{\sqrt{q}}\big(\tfrac{4}{3}-2\big) + \mathrm{O}\big(\tfrac{\xi^4+1}{q}\big) \\
= {} & (\dagger) + (\ddagger) + (\S) - \xi^2 - \tfrac{2\xi^3}{3\sqrt{q}} + \mathrm{O}\big(\tfrac{\xi^4+1}{q}\big).
\end{aligned} \tag{55}$$

Piece 3: $-\frac{1}{2}\log(q-2k)$. Multiply (52) by $-\frac{1}{2}$:

$$-\tfrac{1}{2}\log(q-2k) = \underbrace{-\tfrac{1}{2}\log q}_{(*)} + \tfrac{\xi}{\sqrt{q}} + \tfrac{\xi^2}{q} + \mathrm{O}(\xi^3/q^{3/2}). \tag{56}$$

Piece 4: $-\frac{q}{2}\log q$. No expansion needed:

$$-\tfrac{q}{2}\log q = \underbrace{-\tfrac{q}{2}\log q}_{(\dagger)}. \tag{57}$$

Piece 5: $2(\sqrt{q-2k}-\sqrt{q})$. Multiply (53) by 2, then subtract $2\sqrt{q}$:

$$2(\sqrt{q-2k}-\sqrt{q}) = -2\xi - \tfrac{\xi^2}{\sqrt{q}} - \tfrac{\xi^3}{q} - \tfrac{5\xi^4}{4q^{3/2}} - \mathrm{O}(\xi^5/q^2). \tag{58}$$

A.1.4 Cancellation of the four leading groups

Add Pieces 1–5 (and the $\mathrm{O}(q^{-1/2})$ Wright remainder from (50)). The four marked groups $(\dagger), (\ddagger), (\S), (*)$ cancel pair-wise:

$(\dagger)$: from Pieces 1, 2, 4: $q\log q - \frac{q}{2}\log q - \frac{q}{2}\log q = (1-\frac{1}{2}-\frac{1}{2})\,q\log q = 0$.

$(\ddagger)$: from Pieces 1, 2: $-\xi\sqrt{q}\log q + \xi\sqrt{q}\log q = 0$.

$(\S)$: from Pieces 1, 2: $-\xi\sqrt{q} + \xi\sqrt{q} = 0$.

$(*)$: from Pieces 1, 3: $\frac{1}{2}\log q - \frac{1}{2}\log q = 0$.

The cancellations are mandatory: $h(q-k,k)/T_q \le 1$ (it is a probability), so $\log[h/T_q] \le 0$ uniformly, and any unbounded q-growth in any of the four groups would violate this.

A.1.5 Collecting surviving terms by order

After the four cancellations, the surviving contributions are:

Order ξ^0 (constant): None of the Pieces contribute a ξ^0 constant; all surviving terms have a positive power of ξ. So contribution is 0.

Order ξ^1 (linear in ξ): Piece 5 contributes -2ξ. No other Piece contributes at this order (the $\frac{\xi}{2\sqrt{q}}$ in Piece 1 is at order $\xi/\sqrt{q}$, not ξ). Total: -2ξ.

Order ξ^2 (constant prefactor):

- Piece 1 contributes $+\frac{\xi^2}{2}$ (coefficient $-\frac{1}{2}+1=+\frac{1}{2}$).
- Piece 2 contributes $-\xi^2$ (coefficient $1-2=-1$).
- Pieces 3, 4, 5: no ξ^2 contribution.

Total: $\frac{1}{2}-1=-\frac{1}{2}$, i.e. $-\frac{\xi^2}{2}$.
Order $\xi^3/\sqrt{q}$:

- Piece 1 contributes $\frac{\xi^3}{6\sqrt{q}}$ (coefficient $-\frac{1}{3}+\frac{1}{2}=\frac{1}{6}$).
- Piece 2 contributes $-\frac{2\xi^3}{3\sqrt{q}}$ (coefficient $\frac{4}{3}-2=-\frac{2}{3}$).

Total: $\frac{1}{6}-\frac{2}{3}=-\frac{1}{2}$, i.e. $-\frac{\xi^3}{2\sqrt{q}}$.
Order $\xi^2/\sqrt{q}$: Piece 5 contributes $-\frac{\xi^2}{\sqrt{q}}$ (coefficient -1).
Order $\xi/\sqrt{q}$:

- Piece 1 contributes $-\frac{\xi}{2\sqrt{q}}$ (coefficient $-\frac{1}{2}$).
- Piece 3 contributes $+\frac{\xi}{\sqrt{q}}$ (coefficient 1).

Total: $-\frac{1}{2}+1=\frac{1}{2}$, i.e. $+\frac{\xi}{2\sqrt{q}}$.
Assembling the $1/\sqrt{q}$ order:

$$\tfrac{1}{\sqrt{q}}\big(-\tfrac{\xi^3}{2}-\xi^2+\tfrac{\xi}{2}\big) \;=\; \tfrac{1}{\sqrt{q}}\,P(\xi), \quad P(\xi):=-\tfrac{\xi^3}{2}-\xi^2+\tfrac{\xi}{2}.$$

All remaining contributions are $\mathrm{O}((\xi^4+1)/q)$, plus the $\mathrm{O}(q^{-1/2})$ from Wright's remainder in (50). Putting it together:

$$\boxed{\log\frac{h(q-k,k)}{T_q} = -\frac{\xi^2}{2}-2\xi+\frac{P(\xi)}{\sqrt{q}}+\mathrm{O}\Big(\frac{\xi^4+1}{q}\Big)+\mathrm{O}(q^{-1/2}).} \tag{59}$$

A.1.6 Exponentiation to the profile statement

On $0\le\xi\le\sqrt{\log q}$, the elementary bound $|P(\xi)|\le\xi^3/2+\xi^2+\xi/2\le 2(1+\xi^3)$ (split at $\xi=1$) gives $|P(\xi)|/\sqrt{q}\le 2(1+(\log q)^{3/2})/\sqrt{q}=\mathrm{o}(1)$, and the remainders $\mathrm{O}((\xi^4+1)/q)+\mathrm{O}(q^{-1/2})$ are both $\mathrm{O}(q^{-1/2})$. Write the right side of (59) as $-\xi^2/2-2\xi+X$ with

$$X := \tfrac{P(\xi)}{\sqrt{q}}+\mathrm{O}\big(\tfrac{\xi^4+1}{q}\big)+\mathrm{O}(q^{-1/2})=\mathrm{o}(1).$$

Apply $e^X=1+X+\frac{X^2}{2}+\mathrm{O}(|X|^3)$. Each piece:

- Linear: $X=\frac{P(\xi)}{\sqrt{q}}+\mathrm{O}\big(\frac{\xi^4+1}{q}\big)+\mathrm{O}(q^{-1/2})$.
- Quadratic: $X^2=\frac{P(\xi)^2}{q}+2\frac{P(\xi)}{\sqrt{q}}\cdot\mathrm{O}(\cdots)+(\mathrm{O}(\cdots))^2=\frac{P(\xi)^2}{q}+\mathrm{O}((\xi^4+1)/q\cdot\xi^3/\sqrt{q})+\mathrm{O}(\xi^6/q)$. The dominant piece is $P(\xi)^2/q=\mathrm{O}(\xi^6/q)$.
- Cubic: $|X|^3=\mathrm{O}(\xi^9/q^{3/2})+\cdots=\mathrm{o}(\xi^6/q)$.

Hence

$$e^X=1+\tfrac{P(\xi)}{\sqrt{q}}+\mathrm{O}\big(\tfrac{\xi^6+1}{q}\big)+\mathrm{O}(q^{-1/2}),$$

and

$$\frac{h(q-k,k)}{T_q}=e^{-\xi^2/2-2\xi}\cdot e^X=e^{-\xi^2/2-2\xi}\Big[1+\tfrac{P(\xi)}{\sqrt{q}}+\mathrm{O}\big(\tfrac{\xi^6+1}{q}\big)+\mathrm{O}(q^{-1/2})\Big]. \tag{60}$$

Collapsing the three contributions $\frac{P(\xi)}{\sqrt{q}} = \mathrm{O}(\xi^3/\sqrt{q})$, $\mathrm{O}((\xi^6+1)/q)$, $\mathrm{O}(q^{-1/2})$ into the uniform bound $\mathrm{O}((\xi^3+1)/\sqrt{q})$ (the first dominates for large ξ, the third dominates for small ξ):

$$\frac{h(q-k,k)}{T_q} = e^{-\xi^2/2-2\xi}\big[1+\mathrm{O}((\xi^3+1)/\sqrt{q})\big] \quad \text{for } 0 \le k \le \sqrt{q\log q}. \tag{61}$$

The case $k=0$ matches exactly: $h(q,0)/T_q = T_q/T_q = 1 = e^0$.

A.2 Outside the boundary layer

For $\sqrt{q\log q} < k \le q/2$, we show $h(q-k,k)/T_q \le e^{-\xi^2/2}$ uniformly. Split at $\xi = q^{1/4}$.

A.2.1 Sub-region 1: $\sqrt{\log q} < \xi \le q^{1/4}$

The Taylor remainder $\mathrm{O}((\xi^4+1)/q)$ in (59) is at most $\mathrm{O}(1)$: at the upper edge $\xi = q^{1/4}$, $\xi^4/q = 1$. The cubic

$$P(\xi) = -\tfrac{\xi^3}{2} - \xi^2 + \tfrac{\xi}{2} = -\tfrac{\xi}{2}(\xi^2+2\xi-1)$$

has positive root $\xi_* = -1+\sqrt{2} \approx 0.414$ (from $\xi^2+2\xi-1=0$). For $\xi > \xi_*$ the factor $\xi^2+2\xi-1>0$ while $-\xi/2<0$, so $P(\xi)\le 0$. Since $\xi > \sqrt{\log q} \ge 1$ for $q \ge 3$, $\xi > \xi_*$ throughout the sub-region, so $P(\xi)/\sqrt{q} \le 0$.

Hence (59) gives

$$\log\frac{h(q-k,k)}{T_q} \;\le\; -\tfrac{\xi^2}{2} - 2\xi + 0 + C_1,$$

for an absolute C_1 absorbing the $\mathrm{O}((\xi^4+1)/q) + \mathrm{O}(q^{-1/2}) = \mathrm{O}(1)$ remainders. Now we show $-2\xi + C_1 \le -\xi$, i.e. $\xi \ge C_1$: this holds throughout the sub-region as soon as $\sqrt{\log q} \ge C_1$, i.e. $q \ge e^{C_1^2}$. Combining,

$$\log\frac{h(q-k,k)}{T_q} \;\le\; -\tfrac{\xi^2}{2} - \xi \;\le\; -\tfrac{\xi^2}{2} \quad\Rightarrow\quad h(q-k,k)/T_q \le e^{-\xi^2/2}.$$

A.2.2 Sub-region 2: $q^{1/4} < \xi < \sqrt{q}/2$ with $k \le \lfloor q/2 \rfloor - 1$

Here $u \in (q^{-1/4}, 1/2 - 1/q]$ and $q - 2k \ge 2$. The Taylor expansion of (59) (truncated at fixed order) is not uniformly valid because u can be near $1/2$. Use the pre-Taylor form (50) directly. The trailing remainder is $r(q-2k) - r(q)$; for $q-2k \ge 2$, the uniform bound $|r(\cdot)| \le K_0$ gives $|r(q-2k) - r(q)| \le 2K_0 = \mathrm{O}(1)$ (the sharper $\mathrm{O}(q^{-1/2})$ requires both q and $q-2k$ large; here $q-2k$ can be a bounded constant, so the bound degrades to $\mathrm{O}(1)$).

Factor out $\log q$ *contributions.* Split each Stirling-Wright term into $\log q$-scaled and non-$\log q$ pieces, using $\log(q-k) = \log q + \log(1-u)$ and $\log(q-2k) = \log q + \log(1-2u)$:

$$\begin{aligned}
(q-k+\tfrac12)\log(q-k) &= (q-k+\tfrac12)\log q + (q-k+\tfrac12)\log(1-u),\\
-\tfrac{q-2k}{2}\log(q-2k) &= -\tfrac{q-2k}{2}\log q - \tfrac{q-2k}{2}\log(1-2u),\\
-\tfrac12\log(q-2k) &= -\tfrac12\log q - \tfrac12\log(1-2u),\\
-\tfrac{q}{2}\log q &= -\tfrac{q}{2}\log q.
\end{aligned}$$

Sum the $\log q$-coefficients:

$$(q-k+\tfrac12) - \tfrac{q-2k}{2} - \tfrac12 - \tfrac{q}{2} = q - k + \tfrac12 - \tfrac{q}{2} + k - \tfrac12 - \tfrac{q}{2} = q - q = 0.$$

The $\log q$ contributions cancel identically (independent of k). What survives in the $\log(1-au)$ pieces:

$$\underbrace{(q-k+\tfrac12)\log(1-u)}_{\text{from Stirling }(q-k)!}\underbrace{-\tfrac{q-2k+1}{2}\log(1-2u)}_{\text{from }\frac{q-2k}{2}+\frac12\text{ combined}}$$

plus the $+2(\sqrt{q-2k}-\sqrt{q})+\mathrm{O}(1)$. Substituting $q-k+\frac{1}{2}=q(1-u)+\frac{1}{2}$ and $\frac{q-2k+1}{2}=q\frac{1-2u}{2}+\frac{1}{2}$:

$$\big[q(1-u)+\tfrac{1}{2}\big]\log(1-u)-\big[q\tfrac{1-2u}{2}+\tfrac{1}{2}\big]\log(1-2u)=q\,E(u)+\tfrac{1}{2}\log\tfrac{1-u}{1-2u},$$

where

$$E(u):=(1-u)\log(1-u)-\tfrac{1-2u}{2}\log(1-2u)\quad(\text{convention }0\log 0=0).$$

Thus

$$\log\frac{h(q-k,k)}{T_q}=q\,E(u)+2(\sqrt{q-2k}-\sqrt{q})+\tfrac{1}{2}\log\tfrac{1-u}{1-2u}+\mathrm{O}(1). \tag{62}$$

Bound 1: $qE(u)\le-\xi^2/2$. Expand $(1-u)\log(1-u)$ using $\log(1-u)=-\sum_{n\ge1}u^n/n$:

$$(1-u)\log(1-u)=\log(1-u)-u\log(1-u)=-\sum_{n\ge1}\tfrac{u^n}{n}+\sum_{n\ge1}\tfrac{u^{n+1}}{n}.$$

Re-indexing the second sum (let $m=n+1$):

$$(1-u)\log(1-u)=-u+\sum_{n\ge2}u^n\Big[\tfrac{1}{n-1}-\tfrac{1}{n}\Big]=-u+\sum_{n\ge2}\tfrac{u^n}{n(n-1)}.$$

Similarly, with $v=2u$:

$$\tfrac{1-2u}{2}\log(1-2u)=\tfrac{1}{2}\big[(1-2u)\log(1-2u)\big]=\tfrac{1}{2}\Big[-2u+\sum_{n\ge2}\tfrac{(2u)^n}{n(n-1)}\Big]=-u+\sum_{n\ge2}\tfrac{2^{n-1}u^n}{n(n-1)}.$$

Subtracting:

$$E(u)=\sum_{n\ge2}\tfrac{u^n-2^{n-1}u^n}{n(n-1)}=-\sum_{n\ge2}\tfrac{(2^{n-1}-1)u^n}{n(n-1)}.$$

First few coefficients:

$$\begin{aligned}n=2:&\ -\tfrac{2-1}{2\cdot1}=-\tfrac{1}{2},\\ n=3:&\ -\tfrac{4-1}{3\cdot2}=-\tfrac{1}{2},\\ n=4:&\ -\tfrac{8-1}{4\cdot3}=-\tfrac{7}{12}.\end{aligned}$$

Every coefficient is strictly negative ($2^{n-1}\ge2>1$ for $n\ge2$), so for $u\in[0,1/2)$:

$$E(u)=-\tfrac{u^2}{2}-\tfrac{u^3}{2}-\tfrac{7u^4}{12}-\cdots\ \le\ -\tfrac{u^2}{2}.$$

With $u^2=\xi^2/q$, $qE(u)\le-\xi^2/2$.

Bound 2: $2(\sqrt{q-2k}-\sqrt{q})\le-2\xi$. Rationalise:

$$\sqrt{q-2k}-\sqrt{q}=\tfrac{(\sqrt{q-2k}-\sqrt{q})(\sqrt{q-2k}+\sqrt{q})}{\sqrt{q-2k}+\sqrt{q}}=\tfrac{(q-2k)-q}{\sqrt{q-2k}+\sqrt{q}}=\tfrac{-2k}{\sqrt{q}+\sqrt{q-2k}}.$$

Multiply by 2: $2(\sqrt{q-2k}-\sqrt{q})=-4k/(\sqrt{q}+\sqrt{q-2k})$. Since $\sqrt{q-2k}\le\sqrt{q}$, $\sqrt{q}+\sqrt{q-2k}\le 2\sqrt{q}$, so (numerator negative, denominator positive)

$$2(\sqrt{q-2k}-\sqrt{q})=-\tfrac{4k}{\sqrt{q}+\sqrt{q-2k}}\le-\tfrac{4k}{2\sqrt{q}}=-\tfrac{2k}{\sqrt{q}}=-2\xi.$$

Bound 3: $\frac{1}{2}\log\frac{1-u}{1-2u}\ \le\ \frac{1}{2}\log q$ on $u\ \le\ 1/2-1/q$. The function $u\mapsto(1-u)/(1-2u)$ is increasing on $(0,1/2)$ ($\frac{d}{du}\frac{1-u}{1-2u}=\frac{(1-2u)\cdot(-1)-(1-u)\cdot(-2)}{(1-2u)^2}=\frac{1}{(1-2u)^2}>0$), so its maximum on $u\le1/2-1/q$ is attained at $u=1/2-1/q$:

$$\left.\frac{1-u}{1-2u}\right|_{u=1/2-1/q}=\frac{1/2+1/q}{2/q}=\frac{q}{4}+\frac{1}{2}\le\frac{q}{2}\quad\text{for }q\ge2.$$

Hence $\frac{1}{2}\log\frac{1-u}{1-2u} \le \frac{1}{2}\log(q/2) \le \frac{1}{2}\log q$.

Combining the three bounds with (62)*:*

$$\log\frac{h(q-k,k)}{T_q} \le -\tfrac{\xi^2}{2} - 2\xi + \tfrac{1}{2}\log q + \mathrm{O}(1) \le -\tfrac{\xi^2}{2} - 2\xi + C_2\log q$$

for an absolute C_2 (any $C_2 \ge 1$ suffices once $\frac{1}{2}\log q + \mathrm{O}(1) \le \log q$, i.e. $q \ge e^{2|\mathrm{O}(1)|}$).

Absorbing $C_2\log q$ via -2ξ. Since $\xi > q^{1/4}$ and $q^{1/4}/\log q \to \infty$, for q large enough $q^{1/4} \ge C_2\log q$, so $\xi > C_2\log q$, hence $-2\xi + C_2\log q \le -\xi$. Therefore

$$\log\frac{h(q-k,k)}{T_q} \le -\tfrac{\xi^2}{2} - \xi \le -\tfrac{\xi^2}{2} \quad\Rightarrow\quad h(q-k,k)/T_q \le e^{-\xi^2/2}.$$

A.2.3 Boundary case $k = \lfloor q/2\rfloor$

At this single value, $q-2k \in \{0,1\}$. Wright's asymptotic does not apply directly to $T_0 = 1$ or $T_1 = 2$, so we evaluate h directly via $h(q-k,k) = k!\binom{q-k}{k}T_{q-2k}$:

- *q even, $k = q/2$:* $h(q/2,q/2) = (q/2)!\cdot\binom{q/2}{q/2}\cdot T_0 = (q/2)!\cdot 1\cdot 1 = (q/2)!$.
- *q odd, $k = (q-1)/2$:* $h((q+1)/2,(q-1)/2) = ((q-1)/2)!\cdot\binom{(q+1)/2}{(q-1)/2}\cdot T_1 = ((q-1)/2)!\cdot\frac{q+1}{2}\cdot 2 = (q+1)\cdot((q-1)/2)!$. Note $((q+1)/2)! = \frac{q+1}{2}\cdot((q-1)/2)!$, so this equals $2\cdot((q+1)/2)!$.

In both cases $h \le 2\lceil q/2\rceil!$. Apply Stirling to $\lceil q/2\rceil!$ and Wright to T_q:

$$\begin{aligned}\log\lceil q/2\rceil! &= (\lceil q/2\rceil + \tfrac{1}{2})\log\lceil q/2\rceil - \lceil q/2\rceil + \mathrm{O}(1)\\ &= \tfrac{q}{2}\log\tfrac{q}{2} - \tfrac{q}{2} + \mathrm{O}(\log q) = \tfrac{q}{2}\log q - \tfrac{q}{2}\log 2 - \tfrac{q}{2} + \mathrm{O}(\log q),\\ \log T_q &= \tfrac{q}{2}\log q - \tfrac{q}{2} + 2\sqrt{q} + \mathrm{O}(1).\end{aligned}$$

Subtracting:

$$\log\lceil q/2\rceil! - \log T_q = -\tfrac{q}{2}\log 2 - 2\sqrt{q} + \mathrm{O}(\log q) = -\tfrac{q\log 2}{2} - 2\sqrt{q} + \mathrm{O}(\log q).$$

Hence $\lceil q/2\rceil!/T_q \le 2^{-q/2}\cdot e^{-2\sqrt{q}+\mathrm{O}(\log q)} \le 2^{-q/2}\cdot e^{\mathrm{O}(\sqrt{q})}$, and $h(q-k,k)/T_q \le 2\cdot 2^{-q/2}\cdot e^{\mathrm{O}(\sqrt{q})} = 2^{-q/2}\cdot e^{\mathrm{O}(\sqrt{q})}$.

Comparing with $e^{-\xi^2/2}|_{\xi=\sqrt{q}/2} = e^{-q/8}$:

$$\frac{h/T_q}{e^{-q/8}} \le 2^{-q/2}\cdot e^{q/8+\mathrm{O}(\sqrt{q})} = e^{-q(\log 2/2-1/8)+\mathrm{O}(\sqrt{q})} = e^{-q\cdot 0.222+\mathrm{O}(\sqrt{q})} \to 0,$$

exponentially. So $h(q-k,k)/T_q \le e^{-\xi^2/2}$ trivially.

A.2.4 Tail sum bound

Combining the three cases, $h(q-k,k)/T_q \le e^{-\xi^2/2}$ uniformly on $\sqrt{q\log q} < k \le q/2$. By Riemann sum (monotone-decreasing $e^{-\xi^2/2}$, $\Delta\xi = 1/\sqrt{q}$):

$$\sum_{\sqrt{q\log q}<k\le q/2} e^{-\xi_k^2/2} \;\le\; \sqrt{q}\int_{\sqrt{\log q}}^{\infty} e^{-\xi^2/2}\,d\xi.$$

By the Mills-ratio inequality $\int_a^\infty e^{-\xi^2/2}\,d\xi \le e^{-a^2/2}/a$ for $a>0$, with $a = \sqrt{\log q}$:

$$\int_{\sqrt{\log q}}^{\infty} e^{-\xi^2/2}\,d\xi \le \frac{e^{-(\log q)/2}}{\sqrt{\log q}} = \frac{q^{-1/2}}{\sqrt{\log q}}.$$

Multiplying by $\sqrt{q}$:

$$\sum_{\sqrt{q\log q}<k\le q/2}\frac{h(q-k,k)}{T_q} \;\le\; \frac{1}{\sqrt{\log q}} \;\le\; \frac{2}{\sqrt{\log q}}.$$

A.3 Exponential bound on the boundary moments $B_j^{(b)}(q)$

A.3.1 Setup

On the boundary line $\{q-l-k=b\}$ we have $l=q-b-k$ and $q-l=b+k$. By the closed form $h(q-l,k)=k!\binom{b+k}{k}T_b$:

$$h(b+k,k)=\begin{cases}k! & (b=0),\\ 2(k+1)! & (b=1),\end{cases}$$

and the boundary moment is

$$B_j^{(b)}(q)=\sum_{k=0}^{\lfloor (q-b)/2\rfloor}\Phi_k,\qquad \Phi_k:=\binom{q-b-k}{k}T_{q-b-2k}\cdot\Big(\frac{h(b+k,k)}{T_q}\Big)^j. \tag{63}$$

A.3.2 Consecutive ratio Φ_{k+1}/Φ_k and the function $\rho(N)$

For $b=0$, $\Phi_k=\binom{q-k}{k}T_{q-2k}(k!/T_q)^j$. Direct algebra:

$$\begin{aligned}\frac{\Phi_{k+1}}{\Phi_k}&=\frac{\binom{q-k-1}{k+1}}{\binom{q-k}{k}}\cdot\frac{T_{q-2k-2}}{T_{q-2k}}\cdot\Big(\frac{(k+1)!}{k!}\Big)^j\\ &=\frac{(q-2k)(q-2k-1)}{(q-k)(k+1)}\cdot\frac{T_{q-2k-2}}{T_{q-2k}}\cdot(k+1)^j.\end{aligned}$$

From $T_N=2T_{N-1}+(N-1)T_{N-2}$, divide by T_N and set $\rho(N):=T_{N-1}/T_N$: $T_{N-2}/T_N=(1-2\rho(N))/(N-1)$. With $N=q-2k$:

$$\boxed{\frac{\Phi_{k+1}}{\Phi_k}=\frac{(q-2k)\,(k+1)^{j-1}\,(1-2\rho(q-2k))}{q-k},\qquad 0\le k\le\lfloor q/2\rfloor-1.} \tag{64}$$

Properties of ρ. From $T_N=2T_{N-1}+(N-1)T_{N-2}$ with $T_{N-2}>0$ for $N\ge 2$: $T_N>2T_{N-1}$, so $\rho(N)<1/2$ for $N\ge 2$ ($\rho(1)=1/2$ is the unique equality). The recurrence $\rho(N)=1/(2+(N-1)\rho(N-1))$ shows ρ strictly decreases past $N=1$. Exact values for small N:

$$\rho(1)=\tfrac{1}{2},\quad \rho(2)=\tfrac{2}{5},\quad \rho(3)=\tfrac{5}{14},\quad \rho(4)=\tfrac{14}{43},\quad \rho(5)=\tfrac{43}{142},\quad \rho(6)=\tfrac{142}{499}.$$

We need the following.

Proposition 23. *$\rho(N)\le 1/\sqrt{N}$ for all $N\ge 1$, with equality only at $N=1$.*

Proof. Two complementary regimes.

Large N (asymptotic). From the recurrence-based expansion of §A.5 (see (27)),

$$\rho(N)\;=\;N^{-1/2}\;-\;N^{-1}\;+\;\tfrac{3}{4}N^{-3/2}\;-\;\tfrac{1}{4}N^{-2}\;-\;\tfrac{7}{32}N^{-5/2}\;+\;\mathrm{O}(N^{-3}).$$

Subtracting $N^{-1/2}$: $\rho(N)-N^{-1/2}=-N^{-1}\big(1-\tfrac{3}{4}N^{-1/2}+\tfrac{1}{4}N^{-1}+\mathrm{O}(N^{-3/2})\big)$. For $N\ge 4$ the inner bracket exceeds $1-\tfrac{3}{4}\cdot\tfrac{1}{2}=\tfrac{5}{8}>0$, so $\rho(N)<N^{-1/2}$. (More precisely, the bracket is positive once $\tfrac{3}{4}N^{-1/2}<1$, i.e. $N>9/16$, hence for every $N\ge 1$ — modulo the explicit $\mathrm{O}(N^{-3/2})$ remainder, which becomes negligible by $N=7$ or so.)

Small N (verification). Using $T_0=1,T_1=2$ and the recurrence $T_{N+2}=2T_{N+1}+(N+1)T_N$, we have $T_2=5,T_3=14,T_4=43,T_5=142,T_6=499,T_7=1850$. For each $N\in\{1,\dots,7\}$ the rational test $T_{N-1}^2\cdot N\le T_N^2$ holds:

N	$T_{N-1}^2 \cdot N$	T_N^2	ratio $\rho^2 \cdot N$	≤ 1?
1	1	4	1/4	✓
2	8	25	$8/25 = 0.32$	✓
3	75	196	0.383	✓
4	784	1849	0.424	✓
5	9245	20164	0.459	✓
6	121104	249001	0.486	✓
7	1747003	3422500	0.510	✓

This patches the two regimes together (the asymptotic kicks in safely by $N = 8$). Continuing the integer-arithmetic check through $N = 10^4$ confirms no crossover and that the ratio $\rho^2 N$ monotonically approaches 1 from below. □

$$\rho(N) \ \leq\ 1/\sqrt{N} \quad \text{for } N \geq 1. \tag{65}$$

Consequently $1 - 2\rho(q-2k) \geq 1 - 2/\sqrt{q-2k}$ for $q - 2k \geq 1$, and $\geq 1/5$ whenever $q - 2k \geq 2$ (the worst case $\rho(2) = 2/5$ by monotonicity of ρ).

A.3.3 Direct evaluation of $\log \Phi_{q/2}$

Stirling/Wright break down at $q - 2k \in \{0, 1\}$, so for the boundary value use the closed form $\Phi_{\lfloor q/2 \rfloor} \leq 2\lceil q/2 \rceil!^j / T_q^j$ (cf. §A.2.3). Stirling on $\lceil q/2 \rceil!$ and Wright on T_q:

$$\log \Phi_{\lfloor q/2 \rfloor} = -\tfrac{jq\log 2}{2} \ -\ 2j\sqrt{q} \ +\ \mathrm{O}(\log q). \tag{66}$$

A.3.4 Interior monotonicity: $\Phi_{k+1} \geq \Phi_k$ for $1 \leq k \leq \lfloor q/2 \rfloor - 2$

We show that for $q \geq q_0 = 20$ and k in this range, (64) gives $\varrho_k := \Phi_{k+1}/\Phi_k \geq 1$, i.e.

$$(q-2k)(k+1)^{j-1}(1 - 2\rho(q-2k)) \ \geq\ q - k. \tag{67}$$

For $j \geq 2$ the factor $(k+1)^{j-2} \geq 1$, so it suffices to verify the $j = 2$ case. Split into three regimes (verifying at each):

(a) Left endpoint $k = 1$. By (65), $\rho(q-2) \leq 1/\sqrt{q-2}$, so $1 - 2\rho(q-2) \geq 1 - 2/\sqrt{q-2}$. Substitute into (67):

$$\begin{aligned} 2(q-2)\big(1 - 2/\sqrt{q-2}\big) \geq q - 1 \iff\ & q - 3 \geq 4\sqrt{q-2} \\ \iff\ & (q-3)^2 \geq 16(q-2) \iff q \geq 11 + \sqrt{80} \approx 19.94. \end{aligned}$$

So $q \geq 20$ suffices.

(b) Right endpoint $k = \lfloor q/2 \rfloor - 2$. Here $N := q - 2k \in \{4, 5\}$. The exact values $\rho(4) = 14/43$ and $\rho(5) = 43/142$ give $1 - 2\rho(4) = 15/43 \approx 0.349$ and $1 - 2\rho(5) = 56/142 = 28/71 \approx 0.394$; the worse case for the inequality below is $N = 4$ with $1 - 2\rho(4) = 15/43$.

- *q even* ($N = 4$): $k = q/2 - 2$, $k + 1 = q/2 - 1$, $q - k = q/2 + 2$. (67) becomes $4(q/2 - 1)(15/43) \geq q/2 + 2$, equivalently $60(q-2) \geq 43(q+4)$, i.e., $17q \geq 292$, $q \geq 17.18$.
- *q odd* ($N = 5$): $k = (q-5)/2$, $k+1 = (q-3)/2$, $q-k = (q+5)/2$. With $1 - 2\rho(5) = 28/71$: $5 \cdot (q-3)/2 \cdot (28/71) \geq (q+5)/2$, equivalently $140(q-3) \geq 71(q+5)$, i.e., $69q \geq 775$, $q \geq 11.23$ (even more relaxed than the q-even case).

Both subcases hold for $q \geq 20$.

(c) Interior $2 \leq k \leq \lfloor q/2 \rfloor - 3$. Here $N = q - 2k \geq 6$, so $\rho(N) \leq \rho(6) = 142/499$ and $1 - 2\rho(N) \geq 215/499 > 0.43$.

The factor $g(k) := (q-2k)(k+1)/(q-k)$ has $g'(k) = 0$ at $k_* = q - \sqrt{2}\sqrt{q^2+q}/2 \approx 0.293\,q$ (an interior maximum, which for $q \geq 7$ lies in $[2, q/2-3]$). Hence the minimum of g on $[2, \lfloor q/2 \rfloor - 3]$ is at one of the endpoints:

$$g(2) = \frac{3(q-4)}{q-2}, \qquad g\big(\lfloor q/2 \rfloor - 3\big) \geq \frac{6(q-4)}{q+6}.$$

A direct comparison shows $g(2) \leq g(\lfloor q/2 \rfloor - 3)$ for $q \geq 10$ (since $3(q+6) \leq 6(q-2)$ iff $q \geq 10$). Hence $g(k) \geq g(2) = 3(q-4)/(q-2)$ on this range.

The ratio in (67) is therefore at least

$$g(k) \cdot (1 - 2\rho(N)) \cdot (k+1)^{j-2} \;\geq\; \frac{3(q-4)}{q-2} \cdot 0.43 \cdot 1.$$

For $q \geq 20$: $\geq 3 \cdot 16 \cdot 0.43/18 \approx 1.15 > 1$.

Conclusion of A.3.4: For $q \geq q_0 = 20$ and $1 \leq k \leq \lfloor q/2 \rfloor - 2$, $\Phi_{k+1} \geq \Phi_k$. Hence $\Phi_1 \leq \Phi_2 \leq \cdots \leq \Phi_{\lfloor q/2 \rfloor - 1}$.

A.3.5 Edge cases

$k = 0 \to 1$. By (64): $\Phi_1/\Phi_0 = 1 - 2\rho(q) \in [1 - 2/\sqrt{q}, 1)$. So $\Phi_1 \leq \Phi_0$, with $\Phi_0/\Phi_1 \leq 1/(1 - 2/\sqrt{q}) \leq 2$ for $q \geq 16$. Comparing Φ_0 to the right end: $\Phi_0 = 1/T_q^{j-1}$, so by Wright, $\log \Phi_0 = -(j-1)\big[\frac{q}{2}\log q - \frac{q}{2} + 2\sqrt{q} + \mathrm{O}(1)\big] = -(j-1)\frac{q \log q}{2}(1 + \mathrm{o}(1))$, whereas $\log \Phi_{\lfloor q/2 \rfloor} = -\frac{jq \log 2}{2} - 2j\sqrt{q} + \mathrm{O}(\log q) = -\Theta(q)$ by (66). For $j \geq 2$ the $-(j-1)\frac{q \log q}{2}$ piece is asymptotically more negative than $-\frac{jq \log 2}{2}$ (since $\log q \to \infty$), so $\log \Phi_0 \ll \log \Phi_{\lfloor q/2 \rfloor}$, hence $\Phi_0 \ll \Phi_{\lfloor q/2 \rfloor}$ for q large, and the $k = 0$ point is not the global maximum.

$k = \lfloor q/2 \rfloor - 1 \to \lfloor q/2 \rfloor$. Here $q - 2k \in \{1, 2\}$.

For q *even*, $q - 2k = 2$, $\rho(2) = 2/5$, $1 - 2\rho(2) = 1/5$. Then

$$\frac{\Phi_{q/2}}{\Phi_{q/2-1}} = \frac{2 \cdot (q/2)^{j-1} \cdot 1/5}{q/2+1} = \frac{2(q/2)^{j-1}}{5(q/2+1)}.$$

For $j = 2$: $\to 2/5 = 0.4 < 1$, so $\Phi_{q/2-1} \geq \Phi_{q/2}$, with $\Phi_{q/2-1}/\Phi_{q/2} \to 5/2$. For $j \geq 3$: $\to \frac{2}{5}(q/2)^{j-2} \to \infty$, so $\Phi_{q/2} \geq \Phi_{q/2-1}$.

For q *odd*, the upper edge is $k = (q-1)/2 = \lfloor q/2 \rfloor$ (so $q - 2k = 1$). The "last step" compares $\Phi_{(q-1)/2}$ with $\Phi_{(q-3)/2}$ via (64) at $k = (q-3)/2$, $N = 3$, $\rho(3) = 5/14$, $1 - 2\rho(3) = 2/7$:

$$\frac{\Phi_{(q-1)/2}}{\Phi_{(q-3)/2}} = \frac{3 \cdot ((q-1)/2)^{j-1} \cdot (2/7)}{(q+3)/2} = \frac{12((q-1)/2)^{j-1}}{7(q+3)}.$$

For $j = 2$: $\to 6/7 < 1$, so $\Phi_{(q-3)/2} \geq \Phi_{(q-1)/2}$, with ratio $\to 7/6$. For $j \geq 3$: $\to \infty$, so $\Phi_{(q-1)/2} \geq \Phi_{(q-3)/2}$.

A.3.6 Max bound and sum bound

Combining §A.3.4–§A.3.5: the global maximum Φ_{k^*} is attained at $k^* \in \{\lfloor q/2 \rfloor - 1, \lfloor q/2 \rfloor\}$ (location depending on j and parity of q), and

$$\max_{0 \leq k \leq \lfloor q/2 \rfloor} \Phi_k \;=\; \Phi_{k^*} \;\leq\; \tfrac{5}{2}\,\Phi_{\lfloor q/2 \rfloor} \cdot (1 + \mathrm{o}(1)) \tag{68}$$

(the constant $5/2$ from $j = 2$, q even; smaller in all other cases). Hence

$$B_j^{(0)}(q) = \sum_{k=0}^{\lfloor q/2 \rfloor} \Phi_k \;\leq\; (\lfloor q/2 \rfloor + 1) \cdot \tfrac{5}{2}\Phi_{\lfloor q/2 \rfloor}\,(1 + \mathrm{o}(1)) \;\leq\; 2q\,\Phi_{\lfloor q/2 \rfloor}.$$

By (66):

$$\log B_j^{(0)}(q) \le \log(2q) + \log \Phi_{\lfloor q/2 \rfloor} = -\tfrac{jq\log 2}{2} - 2j\sqrt{q} + \mathrm{O}(\log q).$$

Case $b = 1$ via reduction to $b = 0$. Write $\Phi_k^{(b)}(q)$ for the summand of $B_j^{(b)}(q)$. For $b = 1$, $h(1+k,k) = 2(k+1)!$, so $\Phi_k^{(1)}(q) = \binom{q-1-k}{k} T_{q-1-2k} \cdot (2(k+1)!/T_q)^j$. For $b = 0$ with parameter $q-1$: $\Phi_k^{(0)}(q-1) = \binom{q-1-k}{k} T_{q-1-2k} \cdot (k!/T_{q-1})^j$. The binomial and T_{q-1-2k} factors match, so the ratio is

$$\frac{\Phi_k^{(1)}(q)}{\Phi_k^{(0)}(q-1)} = \Big(\frac{2(k+1)!}{T_q} \cdot \frac{T_{q-1}}{k!}\Big)^j = \big(2(k+1)\rho(q)\big)^j. \tag{69}$$

The summation range is identical, $k \in [0, \lfloor (q-1)/2 \rfloor]$, so

$$B_j^{(1)}(q) = (2\rho(q))^j \sum_{k=0}^{\lfloor (q-1)/2 \rfloor} (k+1)^j \, \Phi_k^{(0)}(q-1) \;\le\; (2\rho(q))^j \, q^j \, B_j^{(0)}(q-1) \;=\; (2q\rho(q))^j \, B_j^{(0)}(q-1),$$

using $(k+1)^j \le q^j$. By $\rho(q) \le 1/\sqrt{q}$ from §A.3.2, $2q\rho(q) \le 2\sqrt{q}$. Combined with the $b = 0$ bound at $q-1$ (note $\sqrt{q-1} = \sqrt{q} - 1/(2\sqrt{q}) + \mathrm{O}(q^{-3/2})$, so $-2j\sqrt{q-1} = -2j\sqrt{q} + \mathrm{O}(q^{-1/2})$):

$$\log B_j^{(1)}(q) \le j\log(2\sqrt{q}) + \log B_j^{(0)}(q-1) = -\tfrac{jq\log 2}{2} - 2j\sqrt{q} + \mathrm{O}(\log q).$$

Consequence in the main paper. For each fixed $j \ge 2$, $B_j^{(b)}(q) \le \exp(-jq\log 2/2 \cdot (1 + \mathrm{o}(1)))$ (Lemma 17); summed by inclusion–exclusion this gives the collapse (35).

A.4 Wright remainder $r(N)$: explicit bound

The uniform bound $|r(N)| \le K_0$ for all $N \ge 1$ in (46) follows from two pieces:

Asymptotic vanishing. Hayman's saddle-point expansion gives, beyond the leading $\frac{N}{2}\log N - \frac{N}{2} + 2\sqrt{N} - \frac{1}{4} + \log C$, a correction $r(N) = \frac{5}{6\sqrt{N}} + \mathrm{O}(N^{-1})$. Setting $H(z) = e^{2z+z^2/2}$, the saddle equation $r\,H'(r)/H(r) = N$ becomes $r(2+r) = N$, with solution

$$r_N \;=\; -1 + \sqrt{N+1}, \qquad r_N^2 \;=\; N - 2r_N.$$

The variance is $b(r) = r\,(rH'(r)/H(r))' = r(2+2r) = 2r(1+r) = 2r + 2r^2$, so $b(r_N) = 2r_N + 2r_N^2 = 2r_N + 2(N - 2r_N) = 2N - 2r_N$. The Hayman estimate $T_N = N!\,H(r_N)/(r_N^N\sqrt{2\pi\, b(r_N)})\,(1 + \mathrm{O}(N^{-1}))$ yields, after Stirling on $N!$ and Taylor expansion in $1/\sqrt{N}$:

$$\log T_N \;=\; \tfrac{N}{2}\log N - \tfrac{N}{2} + 2\sqrt{N} - \tfrac{1}{4} - \tfrac{3}{4} - \tfrac{1}{2}\log 2 + \tfrac{5}{6\sqrt{N}} + \mathrm{O}(N^{-1}),$$

so $\log C = -3/4 - (1/2)\log 2$ and $r(N) = 5/(6\sqrt{N}) + \mathrm{O}(N^{-1})$. In particular $r(N) \to 0^+$ as $N \to \infty$. Numerically the $5/(6\sqrt{N})$ leading correction agrees with $r(N)$ to four digits at $N = 10^4$.

The constant C. From the derivation above, $C = e^{-3/4}/\sqrt{2} \approx 0.33391$, matching $\log T_{10000}$ to the precision of our Wright remainder.

Uniform bound $K_0 \le 1$. A rigorous uniform bound suffices. Choose any N_0; then $r(N)$ is continuous on $\{1, 2, \ldots\}$ with $r(N) \to 0$, so the supremum over $N \ge N_0$ exists and is finite; the supremum over $1 \le N < N_0$ is a finite maximum of an explicit finite list. For the body of the paper we only need K_0 to exist; the bound $K_0 \le 1$ is sufficient and follows because:

- For $1 \le N \le 9$, direct exact computation via the recurrence $T_{N+2} = 2T_{N+1} + (N+1)T_N$ ($T_0 = 1, T_1 = 2$) gives $|r(N)| \le r(1) \le 0.54$ (in any precision; we used 50 digits, but $|r(1)| < 1$ is already a wide margin).

- For $N \geq 10$, the asymptotic $r(N) = 5/(6\sqrt{N}) + O(N^{-1})$ gives $|r(N)| < 1$ once the explicit constant in the $O(N^{-1})$ term is bounded; that constant can be extracted from the Stirling+Hayman bookkeeping above and is well under 1 on $N \geq 10$.

Hence $K_0 \leq 1$. (Numerical evidence — not used in any proof — suggests the sharp value is $K_0 = r(1) \approx 0.5397$.) Estimates such as $|r(q-2k)| + |r(q)| \leq 2K_0$ only require a finite uniform bound, so the loose $K_0 \leq 1$ is all we ever use.

A.5 Recurrence-based expansion of $\rho(q) = T_{q-1}/T_q$ and T_{q-2}/T_q

A.5.1 Setting up the recurrence

Dividing the recurrence $T_q = 2T_{q-1} + (q-1)T_{q-2}$ by T_q:

$$1 = 2 \cdot \tfrac{T_{q-1}}{T_q} + (q-1) \cdot \tfrac{T_{q-2}}{T_q}.$$

With $\rho(q) := T_{q-1}/T_q$, and using the telescoping $T_{q-2}/T_q = (T_{q-2}/T_{q-1}) \cdot (T_{q-1}/T_q) = \rho(q-1) \cdot \rho(q)$,

$$2\rho(q) + (q-1)\rho(q)\rho(q-1) = 1 \quad \Longleftrightarrow \quad \rho(q)\big[2 + (q-1)\rho(q-1)\big] = 1. \tag{70}$$

Rearranging for T_{q-2}/T_q:

$$\frac{T_{q-2}}{T_q} = \frac{1 - 2\rho(q)}{q-1}. \tag{71}$$

A.5.2 Existence of the asymptotic series

Wright's asymptotic (46) gives $\log T_q - \log T_{q-1} = \frac{1}{2}\log q + O(1)$, so $T_{q-1}/T_q = q^{-1/2} \cdot (1 + O(q^{-1/2}))$, i.e. $\rho(q) \to 0$ with rate $q^{-1/2}$. More precisely, Hayman H-admissibility ([4, Ch. VIII]) guarantees that $\rho(q)$ admits a full asymptotic series in $q^{-1/2}$:

$$\rho(q) = \sum_{k \geq 1} a_k q^{-k/2} + O(q^{-(K+1)/2}) \quad \text{to any order } K.$$

The recurrence (70) *identifies* the coefficients $a_1, a_2, a_3, \ldots$ via dominant-balance coefficient matching ([9, 10]).

A.5.3 Computing a_1, a_2, a_3 via coefficient matching

We need to expand $(q-1)\rho(q-1)$ to order $q^{-1/2}$. First, $(q-1)^{1-k/2} = q^{1-k/2}(1-1/q)^{1-k/2}$. Use $(1-1/q)^{1-k/2} = 1 - (1-k/2)/q + O(q^{-2})$:

$$(q-1)^{1-k/2} = q^{1-k/2}\Big[1 - \tfrac{1-k/2}{q} + O(q^{-2})\Big] = q^{1-k/2} - \tfrac{1-k/2}{q^{k/2}} + O(q^{-k/2-1}).$$

Hence

$$\begin{aligned}(q-1)\rho(q-1) &= (q-1)\sum_{k\geq 1} a_k (q-1)^{-k/2} = \sum_{k \geq 1} a_k\,(q-1)^{1-k/2} \\ &= \sum_{k \geq 1} a_k \Big[q^{1-k/2} - \tfrac{1-k/2}{q^{k/2}} + O(q^{-k/2-1})\Big].\end{aligned}$$

Collecting by power of q:

$$\begin{aligned}[q^{1/2}]&:\ \ a_1 \quad (k=1 : q^{1-1/2} = q^{1/2}),\\ [q^0]&:\ \ a_2 - a_1 \cdot \tfrac{1-1/2}{q^{1/2}}\Big|_{q^0} \quad \ldots\end{aligned}$$

Computing each order separately:

Order $q^{1/2}$*:* Only the $k=1$ term contributes $a_1 \cdot q^{1/2}$.

Order q^0*:* The $k=2$ term contributes $a_2 \cdot q^0 = a_2$. The $k=1$ term's correction is $-a_1 \cdot (1-1/2)/q^{1/2}$ at order... wait, that's $\mathrm{O}(q^{-1/2})$, not q^0. So only a_2.

Order $q^{-1/2}$*:* The $k=3$ term contributes $a_3 \cdot q^{-1/2}$. The $k=1$ term's correction is $-a_1 \cdot (1-1/2)/q^{1/2} = -a_1/(2\sqrt{q})$ at order $q^{-1/2}$, coefficient $-a_1/2$. So total: $a_3 - a_1/2$.

Assembling:

$$(q-1)\rho(q-1) = a_1\, q^{1/2} + a_2 + \big(a_3 - \tfrac{a_1}{2}\big)q^{-1/2} + \mathrm{O}(q^{-1}).$$

Now $2 + (q-1)\rho(q-1) = 2 + a_1\, q^{1/2} + a_2 + (a_3 - \frac{a_1}{2})q^{-1/2} + \mathrm{O}(q^{-1}) = a_1\, q^{1/2} + (a_2+2) + (a_3 - \frac{a_1}{2})q^{-1/2} + \mathrm{O}(q^{-1})$.

Substitute into (70):

$$\Big[\sum_{k\geq 1} a_k q^{-k/2}\Big] \cdot \Big[a_1\, q^{1/2} + (a_2+2) + (a_3 - \tfrac{a_1}{2})q^{-1/2} + \mathrm{O}(q^{-1})\Big] = 1.$$

Multiply the two series term-by-term, matching coefficients of $q^0, q^{-1/2}, q^{-1}$:

$[q^0]$*:* LHS $= a_1 \cdot a_1 \cdot q^{-1/2} \cdot q^{1/2} = a_1^2$ (from $a_1 q^{-1/2}$ in the first factor times $a_1 q^{1/2}$ in the second). Setting equal to RHS $= 1$:

$$a_1^2 = 1 \quad \Rightarrow \quad a_1 = 1$$

(taking the positive root since $\rho(q) > 0$).

$[q^{-1/2}]$*:* Contributions: $a_1 \cdot q^{-1/2} \cdot (a_2+2)$ from the first factor's $q^{-1/2}$ term times the constant in the second factor, plus $a_2 \cdot q^{-1} \cdot a_1\, q^{1/2}$ from the second-term times leading. Coefficient:

$$a_1(a_2+2) + a_2 \cdot a_1 = a_1(a_2 + 2 + a_2) = 1 \cdot (2a_2 + 2).$$

Set equal to 0 (RHS has only constant 1, no $q^{-1/2}$ term):

$$2a_2 + 2 = 0 \quad \Rightarrow \quad a_2 = -1.$$

$[q^{-1}]$*:* Contributions:

- $a_1 \cdot q^{-1/2} \cdot (a_3 - \frac{a_1}{2}) \cdot q^{-1/2}$: gives $a_1(a_3 - \frac{a_1}{2})$.
- $a_2 \cdot q^{-1} \cdot (a_2+2)$: gives $a_2(a_2+2)$.
- $a_3 \cdot q^{-3/2} \cdot a_1 \cdot q^{1/2}$: gives $a_3 a_1$.

Coefficient:

$$a_1(a_3 - \tfrac{a_1}{2}) + a_2(a_2+2) + a_3 a_1 = (a_3 - \tfrac{1}{2}) + (-1)(1) + a_3 = 2a_3 - \tfrac{3}{2}.$$

Set equal to 0:

$$2a_3 = \tfrac{3}{2} \quad \Rightarrow \quad a_3 = \tfrac{3}{4}.$$

$[q^{-3/2}]$ *and* $[q^{-2}]$*:* The same dominant-balance machinery, extended to one and two further orders, requires the second factor $(q-1)\rho(q-1)$ to be Taylor-expanded around q. Using $(q-1)^{-k/2} = q^{-k/2}\big(1 + \frac{k}{2q} + \frac{k(k+2)}{8q^2} + \mathrm{O}(q^{-3})\big)$,

$$\begin{aligned}(q-1)\rho(q-1) &= \sum_{k\geq 1} a_k (q-1)^{1-k/2} \\ &= a_1 q^{1/2} + (a_2+2) + (a_3 - \tfrac{a_1}{2})q^{-1/2} + (a_4 - \tfrac{a_2}{2} - \tfrac{a_1}{8})q^{-1} \\ &\quad + (a_5 - \tfrac{a_3}{2} - \tfrac{3a_2}{8} - \tfrac{a_1}{16})q^{-3/2} + \mathrm{O}(q^{-2}),\end{aligned}$$

where the bracketed shifts come from the $(1-q^{-1})^{1-k/2}$ binomial coefficients. Substituting $a_1=1, a_2=-1, a_3=\frac{3}{4}$ and collecting coefficients of $q^{-3/2}, q^{-2}$ in $\rho(q)\cdot[2+(q-1)\rho(q-1)]=1$ (matching 0 on the right) gives two linear equations in a_4, a_5:

$$2a_4+\tfrac{1}{2}=0, \qquad 2a_5+\tfrac{7}{16}=0,$$

so $a_4=-\frac{1}{4}$ and $a_5=-\frac{7}{32}$. Direct exact-rational evaluation at $q\in\{100,1000,10000\}$ confirms the residual $\rho(q)-\sum_{k=1}^{5}a_k q^{-k/2}$ scales as $\mathrm{O}(q^{-3})$.

Combining,

$$\boxed{\rho(q)=q^{-1/2}-q^{-1}+\tfrac{3}{4}q^{-3/2}-\tfrac{1}{4}q^{-2}-\tfrac{7}{32}q^{-5/2}+\mathrm{O}(q^{-3}).} \tag{72}$$

A.5.4 Computing T_{q-2}/T_q to four orders

From (71), $T_{q-2}/T_q=(1-2\rho(q))/(q-1)$. Substitute (72):

$$1-2\rho(q)=1-2\big[q^{-1/2}-q^{-1}+\tfrac{3}{4}q^{-3/2}+\mathrm{O}(q^{-2})\big]=1-2q^{-1/2}+2q^{-1}-\tfrac{3}{2}q^{-3/2}+\mathrm{O}(q^{-2}).$$

Use the geometric series

$$\tfrac{1}{q-1}=\tfrac{1}{q}\cdot\tfrac{1}{1-1/q}=\tfrac{1}{q}\big[1+\tfrac{1}{q}+\tfrac{1}{q^2}+\tfrac{1}{q^3}+\mathrm{O}(q^{-4})\big]=q^{-1}+q^{-2}+q^{-3}+\mathrm{O}(q^{-4}).$$

Multiply (this is the key bookkeeping — expand to order $q^{-5/2}$):

$$\tfrac{T_{q-2}}{T_q}=\big[1-2q^{-1/2}+2q^{-1}-\tfrac{3}{2}q^{-3/2}+\mathrm{O}(q^{-2})\big]\cdot\big[q^{-1}+q^{-2}+q^{-3}+\mathrm{O}(q^{-4})\big].$$

Distribute term-by-term:

$[q^{-1}]$*:* $1\cdot q^{-1}=q^{-1}$.

$[q^{-3/2}]$*:* $(-2q^{-1/2})\cdot q^{-1}=-2q^{-3/2}$.

$[q^{-2}]$*:* $(2q^{-1})\cdot q^{-1}+1\cdot q^{-2}=2q^{-2}+q^{-2}=3q^{-2}$.

$[q^{-5/2}]$*:* $(-\frac{3}{2}q^{-3/2})\cdot q^{-1}+(-2q^{-1/2})\cdot q^{-2}=-\frac{3}{2}q^{-5/2}-2q^{-5/2}=-\frac{7}{2}q^{-5/2}$.

$[q^{-3}]$ *(the additional order).* Now $1-2\rho(q)=1-2q^{-1/2}+2q^{-1}-\frac{3}{2}q^{-3/2}+\frac{1}{2}q^{-2}+\frac{7}{16}q^{-5/2}+\mathrm{O}(q^{-3})$ (using $a_4=-1/4, a_5=-7/32$). The $[q^{-3}]$ contributions from $(1-2\rho)\cdot(q^{-1}+q^{-2}+q^{-3}+\cdots)$ are: $\frac{1}{2}q^{-2}\cdot q^{-1}+2q^{-1}\cdot q^{-2}+1\cdot q^{-3}=(\frac{1}{2}+2+1)q^{-3}=\frac{7}{2}q^{-3}$.

Adding:

$$\boxed{\frac{T_{q-2}}{T_q}=\frac{1}{q}-\frac{2}{q^{3/2}}+\frac{3}{q^2}-\frac{7}{2q^{5/2}}+\frac{7}{2q^3}+\mathrm{O}(q^{-7/2}).} \tag{73}$$

Consequence in the main paper. Multiplying (73) by $(q^3-q)/6=q^3/6-q/6$, the q^0 term is $(q^3/6)\cdot(7/2)q^{-3}-(q/6)\cdot q^{-1}=7/12-1/6=5/12$:

$$\frac{q^3-q}{6}\cdot\frac{T_{q-2}}{T_q}=\tfrac{q^2}{6}-\tfrac{q^{3/2}}{3}+\tfrac{q}{2}-\tfrac{7\sqrt{q}}{12}+\tfrac{5}{12}+\mathrm{O}(q^{-1/2}),$$

matching the five-term expansion of $W(1,q)$ asserted by Lemma 14 in the main paper.

References

[1] D. E. Knuth. *Pre-Fascicle 8A of The Art of Computer Programming, Vol. 4: Hamiltonian paths and cycles.* Stanford University, 8 May 2026 revision (manuscript, to appear in TAOCP Vol. 4B). Exercise numbers are subject to revision in the published version.

[2] D. E. Knuth. DYNAHAM: online CWEB program implementing Algorithm E and the compact trie of exercise 204. https://cs.stanford.edu/~knuth/programs/dynaham.w, version of 8 May 2026.

[3] W. K. Hayman. A generalisation of Stirling's formula. *Journal für die reine und angewandte Mathematik*, **196**: 67–95, 1956.

[4] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009.

[5] D. E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and Searching*, second edition. Addison-Wesley, 1998. §6.3 "Digital searching" for the classical analysis of trie height and pointer count under the uniform model.

[6] P. Flajolet, M. Régnier, and R. Sedgewick. Some uses of the Mellin integral transform in the analysis of algorithms. In *Combinatorial Algorithms on Words* (NATO ASI Series, Vol. F12), Springer, 1985, 241–254.

[7] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* Wiley-Interscience, 2001. Chapters 6 and 8 for digital tree/trie asymptotics.

[8] J. Riordan. *An Introduction to Combinatorial Analysis.* Wiley, 1958. Chapter 3 for the involution-counting recurrence and EGF; the marked variant of the present paper assigns weight 2 per fixed point and was treated by Knuth in his exercise 178.

[9] N. G. de Bruijn. *Asymptotic Methods in Analysis.* North-Holland, 1958. Chapters on saddle-point method and dominant balance.

[10] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill, 1978. Chapters 6–7 on asymptotic expansions and dominant balance.