Closed form and asymptotics for the two quantities in Exercise 7.2.2.3–306(f)

Shisheng Li*

June 6, 2026

Abstract

For Knuth's exercise 7.2.2.3–306(f) [1] we prove the exact closed form $T_r(m) = \frac{1}{2}\big(4^m - (m+3)C_m\big)$ for the total interval count, and for the total gap we prove $T_d(z) = \mathfrak{L}(R)/z$ with $\mathfrak{L}$ the divisor Lambert series and $R = zA^2$ the nonempty-Dyck generating function, giving the finite divisor sum for $T_d(m)$ and the asymptotic $\mathbb{E}[l_t - s_t] = \frac{1}{2}\log m + (\frac{3\gamma}{2} - 1) + \frac{\log m}{3m} + \frac{\gamma+1/72}{m} + O(\log m/m^2)$ by singularity analysis. The gap itself obeys the limiting law $\Pr[d_t = k] \to 1/((k+1)(k+2))$.

1 Setup

A *Dyck path* of half-length m is a sequence of heights $l_0, \dots, l_{2m}$ with $l_0 = l_{2m} = 0$, $l_{t+1} - l_t = \pm 1$, $l_t \geq 0$; write Dyck_{2m} for the set, $|\mathrm{Dyck}_{2m}| = C_m$. Knuth's interval algorithm (b) maintains a stack of intervals with a dummy $(0, \infty)$ at the bottom; at time t it reads off

$$r_t = \#\{\text{stacked intervals}\}, \qquad s_t = \max\{p : (p,q) \text{ stacked},\ p \leq l_t \leq q\}, \qquad d_t := l_t - s_t \geq 0.$$

We study the four totals over uniformly random (P, t),

$$T_x(m) \;=\; \sum_{P \in \mathrm{Dyck}_{2m}} \sum_{t=0}^{2m-1} x_t(P) \qquad (x \in \{r, l, s, d\}), \qquad \mathbb{E}[x_t] = \frac{T_x(m)}{2m\, C_m},$$

with $d_t = l_t - s_t$, hence $T_d = T_l - T_s$. These are exactly the quantities of the interval algorithm of Knuth's exercise 306(b): r_t is the stack size and s_t the highest saved foot covering l_t. Explicitly the stack $[p_0, q_0], \dots, [p_r, q_r]$ keeps $p_0 = 0$, $q_0 = \infty$, $0 = p_0 < \cdots < p_r$; step t reads $r_t = r$ and $s_t = p_k$ for the largest k with $p_k \leq l_t \leq q_k$, and on a backward step ($l_{t+1} = l_t - 1$) updates the stack by: if $l_{t-1} < l_t$, push ($r \leftarrow r{+}1$, $q_r \leftarrow l_t$); then set $p_r \leftarrow l_t$; then if $p_{r-1} = p_r$, merge ($q_{r-1} \leftarrow \max(q_{r-1}, q_r)$, $r \leftarrow r{-}1$). We carry the example $P = UUDDUUUDDD$ ($m = 5$) throughout; running it gives gaps $(d_t)_{t=0}^{9} = (0,1,2,1,0,0,1,3,1,0)$, summing to 9.

Catalan generating functions. $C_m = \frac{1}{m+1}\binom{2m}{m}$, and $A(z) := \sum_{m \geq 0} C_m z^m$ satisfies

$$A = 1 + zA^2, \qquad A(z) = \frac{1 - \sqrt{1-4z}}{2z}. \tag{1}$$

*daizisheng@gmail.com

Set $R := zA^2 = A - 1 = \sum_{m \ge 1} C_m z^m$ (nonempty paths), so $A = 1 + R$. The first-return decomposition $P = U\,Q\,D\,P'$ is the combinatorial content of (1); iterating it underlies every recursion below. We use repeatedly the power coefficients

$$[z^k]A^N = \frac{N}{2k+N}\binom{2k+N}{k}, \qquad [z^k]R^N = [z^k](zA^2)^N = \frac{N}{k}\binom{2k}{k-N} \quad (k \ge 1). \tag{2}$$

The level sum is classical ([2], Exercise 2.3.4.5–5):

$$T_l(m) = 4^m - (2m+1)C_m, \qquad T_l(z) = \frac{1}{1-4z} - 2zA'(z) - A(z). \tag{3}$$

2 The interval count T_r

Theorem 1. *For every $m \ge 1$,*

$$T_r(m) \;=\; \frac{4^m - (m+3)\,C_m}{2}.$$

Proof. Decompose P into primitive arches $P = \Pi_1 \Pi_2 \cdots \Pi_K$. When Π_1 closes it deposits a permanent foot-1 interval (the *carry*) that survives beneath $\Pi_2, \ldots, \Pi_K$ and adds 1 to r_t at each of their $2m - |\Pi_1|$ interior times, while inside each Π_i the algorithm runs as on Π_i standalone. Hence

$$T_r(P) = \sum_{i=1}^{K} T_r(\Pi_i) + \bigl(2m - |\Pi_1|\bigr). \tag{4}$$

For a primitive arch $\Pi = UQD$ one has $r_t(\Pi) = r_{t-1}(Q)$, so $T_r(\Pi) = T_r(Q) + [Q \neq \varnothing]$; summing over arches of half-length k,

$$\sigma_k := \sum_{|\Pi| = 2k} T_r(\Pi) = T_r(k-1) + C_{k-1} \quad (k \ge 2), \qquad \sigma_1 = 0.$$

Summing (4) over $P \in \mathrm{Dyck}_{2m}$ via the first-arch convolution (Π_1 of half-length j, remainder of half-length $m-j$) and the carry defect $\sum_P (2m - |\Pi_1|) = (m-1)C_m$ gives

$$T_r(m) = \sum_{j=1}^{m} \sigma_j\, C_{m+1-j} + (m-1)C_m = \sum_{n=0}^{m-1} T_r(n)\, C_{m-n} + C_{m+1} + (m-3)C_m.$$

In generating functions, with $T(z) = \sum_{m \ge 1} T_r(m) z^m$ and using $A = 1 + zA^2$,

$$T(z) = \frac{zA(z)}{1-4z} - \bigl(A(z) - 1\bigr). \tag{5}$$

Finally $[z^m]\frac{zA}{1-4z} = \sum_{n=0}^{m-1} 4^n C_{m-1-n} = 2 \cdot 4^{m-1} - (2m-1)C_{m-1}$ by (3)'s convolution $\sum_n 4^n C_{N-n} = 2 \cdot 4^N - (2N+1)C_N$, and $(2m-1)C_{m-1} = \frac{m+1}{2} C_m$; subtracting $[z^m](A-1) = C_m$ collapses to the stated form. ∎

Corollary 2. $\mathbb{E}[l_t - 2r_t] = \dfrac{1}{m} - \dfrac{1}{2}$, *since $T_l - 2T_r = (2-m)C_m$ by (3).*

Proposition 3. *As $m \to \infty$,*

$$\mathbb{E}[r_t] = \frac{T_r(m)}{2mC_m} = \frac{\sqrt{\pi m}}{4} - \frac{1}{4} + \frac{9\sqrt{\pi}}{32\sqrt{m}} + O(m^{-1}).$$

Proof. $C_m = \frac{4^m}{m\sqrt{\pi m}}\bigl(1 - \frac{9}{8m} + O(m^{-2})\bigr)$, so $\frac{4^m}{4mC_m} = \frac{\sqrt{\pi m}}{4} + \frac{9\sqrt{\pi}}{32\sqrt{m}} + O(m^{-3/2})$, while $\frac{(m+3)C_m}{4mC_m} = \frac{1}{4} + O(m^{-1})$; subtract. ∎

3 The gap, and its tail counts

The gap $d_t = l_t - s_t$ is pure path geometry: s_t is the highest interval foot covering l_t, and an interval of foot v is created exactly by a down-step from level v. Tracking the most recent down-step from the current level gives:

Lemma 4. *Let μ_t be the minimum height attained since the last down-step from level l_t (and $\mu_t = -1$ if none). Then $d_t = (l_t - \mu_t) - 1$, and for $\ell \geq 1$,*

$$d_t \geq \ell \iff \mu_t \leq l_t - \ell - 1,$$

i.e. the last dip below the current level has depth $\geq \ell + 1$.

Proof. A foot v is installed at a down-step from level v and is never raised, while its ceiling only rises. Since the last down-step from l_t the path fell to its minimum μ_t; the deepest such step, from level $\mu_t + 1$, installed the foot $\mu_t + 1$, and the path has since returned to l_t without going lower, so this is the highest foot covering l_t: $s_t = \mu_t + 1$ (and $s_t = 0$, $\mu_t = -1$, if no down-step from l_t has occurred). Hence $d_t = l_t - s_t = (l_t - \mu_t) - 1$, and the threshold form follows. $\square$

Accordingly we study the *tail counts*

$$h_\ell(m) := \#\{(P,t) : d_t \geq \ell\} \qquad (P \in \mathrm{Dyck}_{2m},\ 0 \leq t < 2m), \qquad T_d(m) = \sum_{\ell \geq 1} h_\ell(m), \tag{6}$$

the last identity being $d_t = \sum_{\ell \geq 1} [d_t \geq \ell]$.

4 The tail identity by generating functions

For a single path write $h_\ell(P) = \#\{t : d_t \geq \ell\}$ and $w_\ell(P) = \#\{t : s_t = 0,\ l_t = \ell\}$ (the *fresh* level-ℓ events, where the cover query returns only the dummy). The first-arch decomposition $P = \Pi_1 P''$ peels a carry $\mathcal{C} = (1, \max \Pi_1)$ — where $\max \Pi$ is the peak height of an arch Π — that lifts s_t from 0 to 1 at each fresh P''-event of level $\leq \max \Pi_1$; reading $\Pi_1 = UQ'D$ by elevation (run on Q' one level up) lowers ℓ. Together,

$$h_\ell(P) = h_\ell(\Pi_1) + h_\ell(P'') - [\max \Pi_1 \geq \ell]\, w_\ell(P''), \qquad w_\ell(P) = w_\ell(\Pi_1) + [\max \Pi_1 < \ell]\, w_\ell(P''), \tag{7}$$

$$w_\ell(UQ'D) = w_{\ell-1}(Q'), \qquad h_\ell(UQ'D) = h_\ell(Q') + w_{\ell-1}(Q'), \tag{8}$$

the elevation (8) because elevation preserves every gap except at fresh events, where it lifts the gap by one. Let $H_\ell = \sum_m h_\ell(m) z^m$ and $W_\ell = \sum_m w_\ell(m) z^m$. Summing (7) over $P = \Pi_1 P''$ (the arches have GF zA; low arches with $\max \Pi \leq \ell - 1$ have GF $S_{\ell-1} := zA_{\leq \ell-2}$, with $A_{\leq k}$ the height-$\leq k$ Dyck GF), and using the arch tail $\sum_{\text{arch}} h_\ell(\Pi) z^{|\Pi|/2} = z(H_\ell + W_{\ell-1})$ from the elevation (8), gives with $1 - zA = 1/A$, $1 - zA^2 = 2 - A$:

$$H_\ell = zA^2(H_\ell + W_{\ell-1}) - A(zA - S_{\ell-1})W_\ell, \qquad W_\ell(1 - S_{\ell-1}) = zA\, W_{\ell-1}, \qquad W_1 = A^2 - A. \tag{9}$$

Proposition 5. *For every $\ell \geq 1$ and $m \geq 1$,*

$$h_\ell(m) = [z^{m+1}]\, \frac{R^{\ell+1}}{1 - R^{\ell+1}}.$$

Proof. Solving (9) for H_ℓ (with $1 - zA^2 = 2 - A$) and using the W-recursion to cancel the $S_{\ell-1}$ terms gives $H_\ell(2 - A) = W_\ell$, i.e. $H_\ell = W_\ell/(2 - A) = W_\ell/(1 - R)$. The denominators $A_{\le k}$ satisfy the continued fraction $A_{\le k} = 1/(1 - zA_{\le k-1})$, $A_{\le 0} = 1$; induction (using $A = 1 + R$, $zA = R/A$) yields the linchpin

$$1 - zA_{\le k} = \frac{1 - R^{k+3}}{A(1 - R^{k+2})} \quad \Longrightarrow \quad 1 - S_{\ell-1} = \frac{1 - R^{\ell+1}}{A(1 - R^{\ell})}. \tag{10}$$

The W-recursion of (9) then solves to $W_\ell = (1-R)R^{\ell+1}/\big(z(1-R^{\ell+1})\big)$ (base $W_1 = A^2 - A$ matches), whence $H_\ell = W_\ell/(1-R) = R^{\ell+1}/\big(z(1-R^{\ell+1})\big)$ and $h_\ell(m) = [z^m]H_\ell = [z^{m+1}]R^{\ell+1}/(1-R^{\ell+1})$. $\square$

5 The exact gap sum

Theorem 6. *With $\mathfrak{L}(x) := \sum_{n\ge 2} \frac{x^n}{1-x^n} = \sum_{N\ge 2}(d(N) - 1)x^N$ the divisor Lambert series (d = number of divisors),*

$$T_d(z) = \frac{\mathfrak{L}(R)}{z}, \tag{11}$$

and for every $m \ge 1$,

$$T_d(m) = \frac{1}{m+1}\sum_{N=1}^{m+1} N\, d(N)\binom{2m+2}{m+1-N} - \frac{1}{2}\binom{2m+2}{m+1}. \tag{12}$$

Proof. Summing Proposition 5 over ℓ and using $T_d = \sum_\ell h_\ell$ from (6),

$$T_d(m) = \sum_{\ell\ge 1}[z^{m+1}]\frac{R^{\ell+1}}{1 - R^{\ell+1}} = [z^{m+1}]\sum_{n\ge 2}\frac{R^n}{1 - R^n} = [z^{m+1}]\mathfrak{L}(R),$$

which is (11). For the finite form, expand each tail geometrically and read coefficients by (2): with $B_N(m) := [z^{m+1}]R^N = \frac{N}{m+1}\binom{2m+2}{m+1-N}$,

$$h_\ell(m) = \sum_{j\ge 1} B_{(\ell+1)j}(m), \qquad T_d(m) = \sum_{\ell\ge 1}\sum_{j\ge 1} B_{(\ell+1)j}(m) = \sum_{N\ge 2}(d(N) - 1)\, B_N(m),$$

since for fixed N the pairs (ℓ, j) with $(\ell + 1)j = N$, $\ell + 1 \ge 2$, number $d(N) - 1$. This is the divisor function *counted*, not extracted. Folding the "-1" by $\sum_{N=1}^{m+1} N\binom{2m+2}{m+1-N} = \frac{m+1}{2}\binom{2m+2}{m+1}$ gives (12). $\square$

The first stratum collapses to a single binomial, $h_1(m) = \binom{2m}{m-1} = mC_m$, so $\Pr[d_t \ge 1] = \frac{1}{2}$; the higher strata stay genuine finite sums whose recurrence order grows with ℓ (the singularities of $R^{\ell+1}/(1-R^{\ell+1})$ are the $(\ell+1)$-th roots of unity). Summing over ℓ makes these dense on $|R| = 1$, a natural boundary of $\mathfrak{L}(R)$; as a D-finite series has only finitely many singularities, T_d is not D-finite although every stratum is. The limiting law follows from $h_\ell(m) \sim \frac{1}{2(\ell+1)}\binom{2m+2}{m+1}$:

$$\Pr[d_t \ge \ell] \to \frac{1}{\ell + 1}, \qquad \Pr[d_t = k] \to \frac{1}{(k+1)(k+2)} \qquad (m \to \infty). \tag{13}$$

6 Asymptotics of the gap sum

The mean is governed by the singularity of $T_d(z) = \mathfrak{L}(R)/z$ at $z = \frac{1}{4}$, where $R \to 1^-$. Set $\delta := -\log R \to 0^+$.

Lemma 7 (Wigert)**.** *As $\delta \to 0^+$,*

$$\sum_{N\geq 1} d(N)\, e^{-N\delta} = \sum_{n\geq 1} \frac{1}{e^{n\delta}-1} = \frac{\gamma + \log(1/\delta)}{\delta} + \frac{1}{4} - \frac{\delta}{144} + O(\delta^3).$$

Proof. Mellin: $\frac{1}{e^{n\delta}-1}$ has transform $\zeta(s)\Gamma(s)$ on n, so the sum equals $\frac{1}{2\pi i}\int_{(c)} \zeta(s)^2\Gamma(s)\delta^{-s}\,ds$, $c > 1$. Shifting left, the double pole at $s = 1$ gives $\delta^{-1}(\log(1/\delta) + \gamma)$, the simple pole at $s = 0$ gives $\zeta(0)^2 = \frac{1}{4}$, and $s = -1$ gives $\zeta(-1)^2(-\delta) = -\delta/144$ (as $\zeta(-1) = -\frac{1}{12}$); the trivial zeros kill $s = -2k$, and $s = -3$ gives $O(\delta^3)$. (See [5, 4, 3].) $\square$

Subtracting the $n = 1$ term $R/(1-R) = (e^{\delta}-1)^{-1} = \delta^{-1} - \frac{1}{2} + \frac{\delta}{12} + O(\delta^3)$ from Lemma 7 (note $\sum_{n\geq 1}\frac{R^n}{1-R^n} = \sum_N d(N)R^N$, $R = e^{-\delta}$):

$$\mathfrak{L}(R) = \frac{\log(1/\delta) + \gamma - 1}{\delta} + \frac{3}{4} - \frac{13\delta}{144} + O(\delta^3). \tag{14}$$

Theorem 8. *As $m \to \infty$,*

$$\mathbb{E}[l_t - s_t] = \frac{T_d(m)}{2mC_m} = \frac{1}{2}\log m + \Big(\frac{3\gamma}{2} - 1\Big) + \frac{\log m}{3m} + \frac{\gamma + 1/72}{m} + O\Big(\frac{\log m}{m^2}\Big).$$

Proof. Put $\varepsilon := 1 - 4z$, $u := \sqrt{\varepsilon}$. Then $R = \frac{1-u}{1+u}$, so $\delta = -\log R = 2\,\mathrm{arctanh}\, u = 2u + \frac{2}{3}u^3 + O(u^5)$, and $1/z = 4(1 + \varepsilon + O(\varepsilon^2))$. Substituting these and $\log(1/\delta) = \frac{1}{2}\log(1/\varepsilon) - \log 2 + O(\varepsilon)$ into $T_d(z) = \mathfrak{L}(R)/z$ via (14) yields the singular expansion in a Δ-domain at $z = \frac{1}{4}$:

$$T_d(z) = \frac{\log\frac{1}{\varepsilon}}{\sqrt{\varepsilon}} + \frac{2(\gamma - 1 - \log 2)}{\sqrt{\varepsilon}} + 3 + \frac{2}{3}\sqrt{\varepsilon}\,\log\tfrac{1}{\varepsilon} + \frac{24\gamma - 24\log 2 - 49}{18}\sqrt{\varepsilon} + O(\varepsilon^{3/2}\log\tfrac{1}{\varepsilon}).$$

By the transfer theorem ([3], Thm. VI.3), with ψ the digamma function,

$$[z^m]\tfrac{\log(1/\varepsilon)}{\sqrt{\varepsilon}} = \tbinom{2m}{m}\big[\psi(m+\tfrac{1}{2}) + \gamma + 2\log 2\big], \quad [z^m]\tfrac{1}{\sqrt{\varepsilon}} = \tbinom{2m}{m}, \quad [z^m]\sqrt{\varepsilon} = -\tfrac{\binom{2m}{m}}{2m-1},$$

$$[z^m]\sqrt{\varepsilon}\log\tfrac{1}{\varepsilon} = \tfrac{\binom{2m}{m}}{2m-1}\big[2 - \psi(m-\tfrac{1}{2}) - \gamma - 2\log 2\big], \qquad [z^m]O(\varepsilon^{3/2}\log) = O\big(4^m\log m/m^{5/2}\big).$$

Dividing by $2mC_m = \frac{2m}{m+1}\binom{2m}{m}$ (so $\binom{2m}{m}/(2mC_m) = \frac{1}{2} + \frac{1}{2m}$ and $\binom{2m}{m}/((2m-1)2mC_m) = \frac{1}{4m} + \frac{3}{8m^2} + \cdots$), and using $\psi(m+\frac{1}{2}) = \log m + \frac{1}{24m^2} + \cdots$, $\psi(m-\frac{1}{2}) = \log m - \frac{1}{m} - \frac{11}{24m^2} + \cdots$, the four terms contribute $\frac{1}{2}\log m + (\frac{3\gamma}{2} - 1) + \frac{\log m + 3\gamma - 2}{2m}$, 0, $-\frac{\log m + \gamma + 2\log 2 - 2}{6m}$, and $\frac{49 - 24\gamma + 24\log 2}{72m}$ (to $O(\log m/m^2)$). The $\log m/m$ parts give $(\frac{1}{2} - \frac{1}{6})\frac{\log m}{m} = \frac{\log m}{3m}$; in the constant-$/m$ part the $\log 2$ cancels and the rationals combine to $\gamma + \frac{1}{72}$. $\square$

Equivalently $\mathbb{E}[l_t - s_t] = \frac{1}{2}H_m + (\gamma - 1) + o(1)$ with H_m the m-th harmonic number: the answer is essentially half a harmonic number, the γ entering through the divisor density $\sum_{N\leq x} d(N) = x\log x + (2\gamma - 1)x + O(\sqrt{x})$.

Small values, and checks. Every formula is confirmed by direct enumeration:

m	1	2	3	4	5	6	7	8
$T_r(m)$	0	3	17	79	344	1454	6047	24903
$T_d(m)$	1	5	22	92	376	1519	6097	24377

For instance $T_d(3) = 22$ is the sum of the gap totals $9, 5, 4, 3, 1$ over the five paths $UUUDDD$, $UUDUDD$, $UUDDUD$, $UDUUDD$, $UDUDUD$ of Dyck_6. The leading term and the constant $\frac{3\gamma}{2} - 1$ of Theorem 8 need *no* complex analysis: Dirichlet's hyperbola method for $\sum_{N\le x} d(N) = x\log x + (2\gamma-1)x + O(\sqrt{x})$ and the local central limit estimate $\binom{2m+2}{m+1-N}/\binom{2m+2}{m+1} \approx e^{-N^2/m}$ for the binomial weights give them with error $O(m^{-1/4})$; the singularity analysis is needed only for the $\frac{\log m}{3m}$ and $\frac{1}{m}$ corrections. Numerically, exact $T_d(4096)$ gives $\mathbb{E}[l_t - s_t] - \frac{1}{2}\log m = -0.13336$, matching $(\frac{3\gamma}{2} - 1) + \frac{\log m}{3m} + \frac{\gamma+1/72}{m} = -0.13418 + 0.00068 + 0.00014$ to all places shown.

7 Full reuse occurs exactly half the time

Proposition 9. *For every $m \ge 1$, $\#\{(P,t) : s_t = l_t\} = mC_m = \binom{2m}{m-1}$, i.e.* $\Pr[d_t = 0] = \frac{1}{2}$.

Proof. By Lemma 4, $s_t = l_t$ iff the last dip below l_t has depth exactly 1 (or $l_t = 0$). The level-0 times have GF $A^2 - A$; for $v \ge 1$ the configuration (path up to v, a depth-1 descent, an above-v excursion, then down to 0) has GF $\sum_{v\ge1} z^{v+1}A^{2v+3} = \frac{z^2A^5}{1-zA^2} = \frac{z^2A^5}{2-A}$. The total is $(A^2-A)+\frac{z^2A^5}{2-A} = \frac{A(A-1)}{2-A} = zA'(z) = \sum_{m\ge1} mC_m z^m$, and $mC_m = \binom{2m}{m-1}$. □

This matches $h_1(m) = \binom{2m}{m-1}$ of §5: positive gap and zero gap each occur with probability $\frac{1}{2}$.

Two questions remain open: a *non-uniform* bijective proof of the gap-tail identity $h_\ell(m) = [z^{m+1}]R^{\ell+1}/(1-R^{\ell+1})$, and the variance of $l_t - s_t$ (its limiting law $\Pr[d_t = k] \to 1/((k+1)(k+2))$ is (13)).

References

[1] D. E. Knuth, *The Art of Computer Programming*, Vol. 4, Pre-Fascicle 7A, §7.2.2.3 (Constraint Satisfaction), exercise 306. Addison–Wesley.

[2] D. E. Knuth, *The Art of Computer Programming*, Vol. 1, §2.3.4.5, exercise 5 (Dyck-path area). Addison–Wesley.

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

[4] P. Flajolet, X. Gourdon, and P. Dumas, Mellin transforms and asymptotics: harmonic sums, *Theoret. Comput. Sci.* **144** (1995), 3–58.

[5] S. Wigert, Sur la série de Lambert et son application à la théorie des nombres, *Acta Math.* **41** (1916), 197–218.