CERIAS Tech Report 2002-09

The Height of a Binary Search Tree: The Limiting Distribution Perspective

Charles Knessl[1], Wojciech Szpankowski [2]

Center for Education and Research in
Information Assurance and Security
&
[2] Purdue University Department of Computer Science
Purdue University, West Lafayette IN 47907
[1]Dept. of Mathematics, Statistics and Computer Science
University of Illinois at Chicago

The Height of a Binary Search Tree: The Limiting Distribution Perspective

June 9, 2001

Charles Knessl*
Dept. Mathematics, Statistics & Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607-7045
U.S.A
knessl@uic.edu

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

We study the height of the binary search tree – the most fundamental data structure used for searching. We assume that the binary search tree is built from a random permutation of n elements. Under this assumption, we study the limiting distribution of the height as $n \to \infty$. We show that the distribution has six asymptotic regions (scales). These correspond to different ranges of k and n where $\Pr\{\mathcal{H}_n \leq k\}$ is the height distribution. In the critical region (the so–called central region), where most of the probability mass is concentrated, the limiting distribution satisfies a non-linear integral equation. While we cannot solve this equation exactly, we show that both tails of the distribution are roughly of a double exponential form. From our analysis we conclude that the average height $\mathbf{E}[\mathcal{H}_n] \sim A \log n - \frac{3}{2}\frac{A}{A-1} \log\log n$ where $A = 4.311\ldots$ is the unique solution of $x \log x - x - x\log 2 + 1 = 0$, $x > 1$, while the variance $\mathbf{Var}[\mathcal{H}_n] = O(1)$. The second term in the expansion of $\mathbf{E}[\mathcal{H}_n]$ and the rate of growth of the variance were also recently obtained by B. Reed who used probabilistic arguments, while M. Drmota established the growth of the variance by analytic methods. Our analysis makes certain assumptions about the forms of some asymptotic expansions, as well as their asymptotic matching.

Key Words: Binary search trees, limiting height distribution, saddle point method, matched asymptotics, linearization, WKB method, non-linear integral equation.

*This work was supported by NSF Grant DMS 99-71656.

†The work of this author was supported by NSF Grants NCR-9415491 and CCR-9804760, and contract 1419991431A from sponsors of CERIAS at Purdue.

1 Introduction

A *binary search tree* is a fundamental data structure used for searching and sorting. It can be defined as follows: A root node is created for the first element. Then subsequent elements are directed to the left or right subtree according to whether they are less or greater than the element stored in the root. A consequence of this construction is that the left subtree and the right subtree of the root are binary search trees themselves. The popularity of binary search trees stems from the fact that many sorting algorithms (e.g., Quicksort) can be conveniently represented by them.

To justify the performance of algorithms that are based on the binary search tree, a body of theory was built. For a tree storing n elements it is known that the worst search time is $O(n)$, however, on average it is only $O(\log n)$. To state precisely the average case performance we must introduce a probabilistic model. We adopt here the standard assumption that all $n!$ permutations of the n elements are equally likely. For such a model, we shall analyze here the height $\mathcal{H}_n$ of the binary search tree, that is, the longest path in a randomly built tree.

The analysis of the height of the binary search tree is a challenging problem. There are still some open questions regarding the height. In 1986 Devroye [5] proved that the average height $\mathbf{E}[\mathcal{H}_n]$ satisfies $\mathbf{E}[\mathcal{H}_n] \sim A \log n$ as $n \to \infty$ where $A = 4.31107\ldots$ is a unique solution of $x \log x - x - x \log 2 + 1 = 0$ for $x > 1$. Earlier Pittel [23] had shown that $\mathcal{H}_n / \log n \to \alpha$ almost surely where $\alpha \leq A$. Then, Devroye and Reed [7] established a stronger result, namely that

$$\mathbf{E}[\mathcal{H}_n] = A \log n + O(\log\log n).$$

They also show that the variance $\mathbf{Var}[\mathcal{H}_n] = O((\log\log n)^2)$. However, Robson [25] has shown experimentally that $\mathbf{E}[|\mathcal{H}_n - \mathbf{E}[\mathcal{H}_n]|] = O(1)$ which would suggest that $\mathbf{Var}[\mathcal{H}_n] = O(1)$. This conjecture was recently proved by Reed [24] and Drmota [8, 9]. Reed used probabilistic arguments while Drmota applied analytic tools. To the best of our knowledge there are no results concerning the limiting distribution of $\mathcal{H}_n$.

In this paper we study the limiting distribution of the height $\mathcal{H}_n$. In the course of our analysis we re-establish recent results concerning the average and the variance of the height. We first observe that $\mathcal{H}_{n+1} = \max\{\mathcal{H}_\ell^L, \mathcal{H}_{n-\ell}^R\} + 1$ where $\mathcal{H}_\ell^L$ and $\mathcal{H}_{n-\ell}^R$ are the left and right subtrees. In the above, ℓ is selected with probability $1/(n+1)$ according to the assumed probabilistic model. In view of this, the distribution

$$L_n^k = \Pr\{\mathcal{H}_n \leq k\}$$

satisfies the following recurrence

$$L_{n+1}^{k+1} = \frac{1}{n+1} \sum_{\ell=0}^{n} L_\ell^k L_{n-\ell}^k$$

with $L_0^0 = 1$ and $L_n^0 = 0$ for $n \geq 1$. This is a non-linear recurrence that we solve asymptotically to obtain the limiting distribution of the height.

We show that there are six scales of the distribution L_n^k. These are defined precisely in the next section and correspond to ranges of k where the structure of L_n^k is different for $n \to \infty$. The most important scale is what we call the *central regime* where the distribution undergoes a transition from being close to zero to being close to one. For this scale we derive a non-linear integral equation that we only know how to solve numerically. But we also show that both tails of the asymptotic distribution are of a double exponential form. We also establish

$$\begin{aligned}
\mathbf{E}[\mathcal{H}_n] - A \log n &\sim -\frac{3}{2}\frac{A}{A-1} \log\log n, \\
\mathbf{Var}[\mathcal{H}_n] &= O(1),
\end{aligned}$$

where A is defined above. The second term of the expansion of $\mathbf{E}[\mathcal{H}_n]$ agrees with Reed [24].

These results can be compared to our recent results [15, 16] concerning the limiting distributions of the height in PATRICIA tries and in digital search trees. The recurrences considered there are

$$\begin{aligned}
h_n^{k+1} &= 2^{-n+1} h_n^{k+1} + 2^{-n} \sum_{i=1}^{n-1} \binom{n}{i} h_i^k h_{n-i}^k, \quad k \geq 0 \\
\bar{h}_{n+1}^{k+1} &= \sum_{i=0}^{n} \binom{n}{i} 2^{-n} \bar{h}_i^k \bar{h}_{h-i}^k, \quad k \geq 0
\end{aligned}$$

where $h_n^k = \mathsf{Pr}\{\mathcal{H}_n^{\mathrm{PAT}} \leq k\}$ and $\bar{h}_n^k = \mathsf{Pr}\{\mathcal{H}_n^{\mathrm{DST}} \leq k\}$ are height distributions for PATRICIA tries and digital search trees, respectively. For these problems, we proved that in the central regime both distributions are spanned on one or two points. This should be compared to the binary search tree where the distribution is spread out over an infinite number of points.

We derive our results using methods of applied mathematics, such as matched asymptotics and the WKB method. These are analytic methods and are especially suitable for problems that cannot be solved exactly by transform methods. They make certain assumptions about the forms of asymptotic expansions (e.g., see (4.4) or (7.5)), and also the asymptotic matching between various scales. We also applied other analytic tools such as linearization, asymptotic matching, the Euler-Maclaurin summation formula, and methods for solving integral, PDE, and recurrence equations.

The paper is organized as follows. In the next section, we present our main results for binary search trees (cf. Theorem 1). The derivations of these results are relegated to Sections 3–8. In Sections 9-10 we present detailed numerical results and discuss consequences of our findings.

2 Summary of Main Results

We let $\mathcal{H}_n$ be the height in a binary search tree and denote the probability distribution by

$$L_n^k = \mathsf{Pr}\{\mathcal{H}_n \leq k\}; \qquad n, k \geq 0. \tag{2.1}$$

As mentioned above, it satisfies the non-linear recursion equation

$$L_{n+1}^{k+1} = \frac{1}{n+1} \sum_{\ell=0}^{n} L_\ell^k L_{n-\ell}^k, \qquad k \geq 0 \tag{2.2}$$

subject to the initial condition

$$L_n^0 = \delta_{0n} \tag{2.3}$$

where δ_{0n} is the Kronecker delta. From (2.2) and (2.3) we can easily show that $L_n^k = 0$ for $n \geq 2^k$ and $L_n^k = 1$ for $n \leq k$. Indeed, the height in a complete binary search tree is at least $\log_2 n$ while the height of a degenerate binary search tree is n. It therefore suffices to consider the range $k < n < 2^k$ (or $\log_2 n < k < n$).

By introducing the generating function

$$G_k(x) = \sum_{n=0}^{\infty} x^n L_n^k$$

we find that it satisfies

$$G'_{k+1}(x) = [G_k(x)]^2, \qquad G_k(0) = 1. \tag{2.4}$$

It follows that $G_k(x)$ is a polynomial in x of degree $2^k - 1$, and the first $k+1$ coefficients in the polynomial are 1. Below we give the first few $G_k(x)$:

$$G_0(x) = 1, \quad G_1(x) = 1 + x, \quad G_2(x) = 1 + x + x^2 + \tfrac{1}{3}x^3,$$
$$G_3(x) = 1 + x + x^2 + x^3 + \tfrac{2}{3}x^4 + \tfrac{1}{3}x^5 + \tfrac{1}{9}x^6 + \tfrac{1}{63}x^7.$$

It is not difficult to solve (2.2) and (2.3) explicitly if n is close to either k or 2^k. In Section 5 we show that

$$\begin{aligned}
k = n-1: \quad & L_n^k = 1 - \tfrac{2^{n-1}}{n!} \qquad (n \geq 1), \\
k = n-2: \quad & L_n^k = 1 - \tfrac{2^{n-3}}{(n-2)!} \qquad (n \geq 3), \\
k = n-3: \quad & L_n^k = 1 - \tfrac{2^{n-6}}{n!} \left[n(n-1)(n-2)\left(n - \tfrac{1}{3}\right) - 40 \right] \qquad (n \geq 5).
\end{aligned}$$

We also have $L_2^0 = 0$ and $L_5^2 = L_4^1 = L_3^0 = 0$. In Section 3 we consider n close to 2^k and show that

$$\begin{aligned}
n = 2^k - 1: \quad & L_n^k = 2^{k+2-3\cdot 2^{k-1}} \left[\prod_{i=2}^{k} (1-2^{-i})^{2^{-i}} \right]^{-2^k} \equiv \Delta_k, \\
n = 2^k - 2: \quad & L_n^k = (2^k - 1)\Delta_k.
\end{aligned}$$

It appears difficult to solve (2.2) (or (2.4)) exactly and obtain an explicit expression for L_n^k. Therefore we consider the asymptotic limit $n \to \infty$. The structure of the problem depends on the relative size of k and n. As with any problem in asymptotic analysis, it is important to identify the basic scales. We find that as $n \to \infty$ there are six major ranges of k that lead to different asymptotic behaviors. We list them below:

(i) far left tail, $2^k - n = M = O(1)$, $M \geq 1$,

(ii) left tail, $n 2^{-k} = \omega$ fixed, $0 < \omega < 1$,

(iii) central region, $k = A \log n + B \log\log n + \zeta$ where $\zeta = O(1)$ and $A = 4.311\ldots$ is the unique solution to

$$x \log x - x - x \log 2 + 1 = 0, \qquad x > 1$$

while we shall argue that

$$B = -\frac{3}{2}\frac{A}{A-1}.$$

(iv) near right tail, $k/\log n = \nu$ fixed, $A < \nu$,

(v) right tail, k/n fixed such that $0 < k/n < 1$,

(vi) far right tail, $j = n - k = O(1)$, $j \geq 1$.

Note that $k \sim \log_2 n$ for case (i) and $k - \log_2 n = O(1)$ for case (ii). If we plot the distribution L_n^k as a function of k for a fixed (large) n, then as k increases we move from region (i) to (vi).

We obtain results for L_n^k in the indicated ranges. In some cases we obtain the asymptotic expansion (or at least the leading term) completely, while in other cases we obtain partial information only, which we supplement with numerical studies. The derivation of these results is presented in the next six sections, where we make certain assumptions about the forms of the asymptotic expansions, as well as the asymptotic matching between the various scales. Our main analytic methods are those of linearization, the WKB method, and matched asymptotic expansions (cf. [10, 20, 21]). The WKB method postulates that the solution $F_n(\omega)$ to an equation (e.g., recurrence, functional or differential equation) has the following form as $n \to \infty$

$$F_n(\omega) = n^\gamma e^{-n\phi(\omega)} \left(A(\omega) + \frac{1}{n} A_1(\omega) + \frac{1}{n^2} A_2(\omega) + \cdots \right), \tag{2.5}$$

where $\phi(\omega),\ \gamma,\ A(\omega),\ A_1(\omega), A_2(\omega), \ldots$ are unknown functions that must be determined from the equation itself. Here is what Fedoryuk [10] has to say about such approximations: "*... It is necessary first of all to guess (and no other word will do) in what form to search for the asymptotic form. Of course, this stage – guessing the form of the asymptotic form – is not subject to any formalization. Analogy, experiments, numerical simulation, physical considerations, intuitions, random guesswork; these are the arsenal of means used by any research worker*".

Theorem 1 *Let $\mathcal{H}_n$ be the height of a binary search tree built from a random permutation of n elements. Under the WKB assumption (2.5), the distribution $L_n^k = \mathsf{Pr}\{\mathcal{H}_n \le k\}$ of the height has the following asymptotic expansions:*

(i) FAR LEFT TAIL*:* $2^k - n = M = O(1)$

$$L_n^k \sim \frac{2^{kM}}{(M-1)!} 4 \left(\frac{1}{2\sqrt{2}K_0} \right)^{2^k} \sim \frac{n^M}{(M-1)!} 4 e^{-c_*(M+n)}, \qquad M \ge 1,$$

where

$$\begin{aligned} K_0 &= \prod_{i=2}^{\infty} (1 - 2^{-i})^{2^{-i}} = .9103\ldots, \\ c_* &= \frac{3}{2}\log 2 + \log(K_0) \end{aligned}$$

(ii) LEFT TAIL*:* $n2^{-k} = \omega, \qquad 0 < \omega < 1$

$$\begin{aligned} L_n^k &\sim \sqrt{n} A(\omega) e^{-n\phi(\omega)} \\ A(\omega) &= 2\sqrt{\frac{2\omega}{\pi}} \sqrt{\omega\phi''(\omega) + 2\phi'(\omega)}\, e^{-(\phi(\omega)+\omega\phi'(\omega))}. \end{aligned}$$

The function $\phi(\omega)$ is calculated numerically in Section 9. Asymptotically we have

$$\begin{aligned} \phi(\omega) &\sim c_* + (1-\omega)\log(1-\omega) + (1-\omega)(c_* - 1), \qquad \omega \uparrow 1 \\ \phi(\omega) &\sim c\omega^{\frac{1}{A\log 2 - 1}} \left(\frac{-\log\omega}{A\log 2 - 1} \right)^{\frac{B\log 2}{A\log 2 - 1}}, \qquad \omega \downarrow 0, \end{aligned}$$

where c is a constant.

(iii) CENTRAL REGION*:* $\zeta = k - A\log n - B\log\log n$

$$L_n^k \sim f(\zeta),$$

where $f(\zeta)$ satisfies the non-linear integral equation

$$f(\zeta+1) = \int_0^1 f(\zeta - A\log x) f(\zeta - A\log(1-x))dx, \qquad -\infty < \zeta < \infty.$$

Asymptotically we have

$$\begin{aligned} f(\zeta) &\sim 1 - c_1\zeta\exp\left(-\left(1-\frac{1}{A}\right)\zeta\right), \qquad \zeta\to+\infty \\ f(\zeta) &\sim 2\sqrt{\frac{2c}{\pi}}\sqrt{\frac{A\log 2}{A\log 2-1}}e^{-\beta\zeta/2}\exp(-ce^{-\beta\zeta}), \qquad \zeta\to-\infty, \end{aligned}$$

where c_1 is a constant,

$$\beta = \frac{\log 2}{A\log 2 - 1} = .3486\ldots,$$

and the constant c is the same constant as in (ii)*.*

(iv) NEAR RIGHT TAIL*:* $k = \nu\log n$, $A < \nu < \infty$

$$1 - L_n^k \sim e^{-a(\nu)\log n}(\log n)^{-1/2}b(\nu),$$

where

$$a(\nu) = \nu\log\nu - \nu - \nu\log 2 + 1,$$

and asymptotically $b(\nu)$ satisfies

$$\begin{aligned} b(\nu) &\sim \frac{1}{2\pi}e^{\nu}\nu^{-\nu-1}, \qquad \nu\to\infty \\ b(\nu) &\sim c_1(\nu - A), \qquad \nu\to A. \end{aligned}$$

(v) RIGHT TAIL*:* $k = n - j$, $\alpha = n/j$, $1 < \alpha < \infty$

$$1 - L_n^k \sim \frac{(2e)^k n^{-k}}{2\pi n^2} z_*^{k-n}(1-z_*)^{-k}\frac{\sqrt{\alpha}z_*}{\sqrt{1-z_*}\sqrt{\alpha z_* - 1}}$$

where $z_* = z_*(\alpha)$ is the unique solution of

$$\frac{1}{\alpha} = \sum_{m=1}^{\infty}\frac{z_*^m}{m(m+1)} = 1 + \frac{(1-z_*)\log(1-z_*)}{z_*}, \qquad 0 < z_* < 1.$$

(vi) FAR RIGHT TAIL*:* $k = n - j$, $j = O(1)$, $j \geq 1$

$$1 - L_n^k \sim \frac{2^n}{n!}\frac{n^{2j-2}}{(j-1)!}2^{1-2j}$$

for $n\to\infty$.

Remark. In the derivation of the above results, we used the WKB method several times. In particular, the analysis for case (ii) assumes the WKB form (4.4) of Section 4, and the conclusions about the behavior of $\phi(\omega)$ as $\omega \uparrow 1$ and $\omega \downarrow 0$ are based on the asymptotic matching between cases (ii) and (i) and cases (ii) and (iii), respectively. The analysis of (iii) assumes the form (8.3), of Section 8, that of (iv) assumes the WKB form (7.5) of Section 7, while case (v) assumes relation below (6.8) (i.e., that of $F_n^j \sim \widetilde{F}_n^j$). □

We observe that in cases (i), (v) and (vi) we have completely determined the leading term. In case (ii) we do not have an exact expression for the function $\phi(\omega)$, but it is relatively easy to compute numerically, as discussed in Section 9. The most difficult cases seem to be (iii) and (iv). In the former we must numerically solve the non-linear integral equation and in the latter we have the unknown function $b(\nu)$. Our analysis suggests the following.

Corollary 1 *The mean of $\mathcal{H}_n$ behaves as*

$$\mathbf{E}[\mathcal{H}_n] - A \log n \sim B \log \log n \tag{2.6}$$

where

$$B = -\frac{3}{2} \frac{A}{A-1}. \tag{2.7}$$

The variance is

$$\mathbf{Var}[\mathcal{H}_n] = O(1) \tag{2.8}$$

for $n \to \infty$.

The above value of B is deduced on the assumption of asymptotic matching between the ζ and ν scales. This implies that $b(\nu)$ vanishes at $\nu = A$ with $b'(A) = c_1 > 0$. The leading term $\mathbf{E}[\mathcal{H}_n] \sim A \log n$ is well-known, having been established by a variety of approaches, both analytical and probabilistic (cf. [5, 7, 8, 9]). It was also conjectured that

$$B = -\frac{1}{2} \frac{A}{A-1}. \tag{2.9}$$

While our value in (2.7) does make certain assumptions about the forms of various asymptotic expansions and their asymptotic matching, we can show that if (2.9) were true then $b(\nu)$ would not vanish at $\nu = A$, and then the solution to the integral equation for $f(\zeta)$ would become negative for certain values of ζ (see Section 8 and Appendix C). This contradiction would seem to exclude the possibility of (2.9). We also do some numerical studies in Section 10 which test the conjectures (2.7) and (2.9). The correct value of B was also established by Reed [24].

Finally, the result for the central region implies that the variance $\mathbf{Var}[\mathcal{H}_n] = O(1)$. This was conjectured by Robson [25] but eluded analysis for some time. Devroye and Reed proved in [7] that $\mathbf{Var}[\mathcal{H}_n] = O(\log\log n)$, and only recently Reed [24] and Drmota [9] proved that $\mathbf{Var}[\mathcal{H}_n] = O(1)$.

From our results we see that in the right tail the rough order of magnitude of $1 - L_n^k$ is $O(e^{-n\log n})$, $O(e^{-k\log n})$, and $O(n^{-a(\nu)})$, for the j, α, and ν scales, respectively. In the near right tail $1 - L_n^k$ is only algebraically small. In the left tail we have $L_n^k = O(e^{-n\phi(\omega)})$. It is also interesting to compare the present results to corresponding ones we previously obtained for digital trees, such as PATRICIA tries [15] and digital search trees (in short: DST) [16]. In these other models the M, ω, α and j scales also arose, and their analysis was somewhat similar to that here. However, the central and near right tail regimes did not occur in the PATRICIA and DST models. Here the limit $n \to \infty$ made the probability mass concentrate at one or two values of k. However, for the BST model, it is spread out over an infinite number of points. Also, the PATRICIA and DST models had certain oscillations occurring in the range of k where the probability mass concentrates, and these seem to be completely absent in the BST model.

From (ii) we see that the function $\phi(\omega)$ is finite at $\omega = 1$, but its derivative has a logarithmic singularity there. Since $1/(A\log 2 - 1) = .5029\ldots$, the function vanishes at $\omega = 0$ but its derivative is infinite. We contrast this to the PATRICIA and DST models, where the corresponding $\phi(\cdot)$ and *all* its derivatives vanished as $\omega \to 0$. We also note that the solution to the integral equation for $f(\zeta)$ is not unique: if $f_0(\zeta)$ is a solution so is $f_0(\zeta + c)$ for any c. To uniquely specify the solution we need the behavior of $f(\zeta)$ as $\zeta \to \infty$ (or $\zeta \to -\infty$). Note also that $f_0(\zeta) = 0$ and $f_0(\zeta) = 1$ are solutions, but these do not have the appropriate behaviors as $\zeta \to \pm\infty$.

Finally, we show in Section 8 that if we set

$$f(\zeta) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \exp\left(\frac{\eta}{A}e^{-\zeta/A}\right) F(\eta)d\eta \tag{2.10}$$

where $a > 0$, then $F(\eta)$ satisfies

$$-F'(\eta) = e^{-2/A}[F(\eta e^{-1/A})]^2. \tag{2.11}$$

The above is precisely the retarded differential equation studied by Drmota [8]. Its solution is also not unique: if $F_0(s)$ is a solution, so is $\lambda F_0(\lambda s)$ for any λ. In Section 8 we discuss the solution to (2.11), using the normalization $F(0) = 1$, which is also used in [8].

3 Far Left Tail

We consider $2^k - n = M = O(1)$, which is close to the left boundary of the support of the probability distribution (we recall that $L_n^k = 0$ for $n \geq 2^k$). We isolate the leading coefficients in the polynomial $G_k(x)$ by writing

$$G_k(x) = c_k x^{2^k-1} + d_k x^{2^k-2} + e_k x^{2^k-3} + \cdots \, . \tag{3.1}$$

It follows that

$$G_k^2(x) = c_k^2 x^{2^{k+1}-2} + 2c_k d_k x^{2^{k+1}-3} + (2c_k e_k + d_k^2) x^{2^{k+1}-4} + \cdots$$

By comparing the above to $G'_{k+1}(x)$ we are led to the recurrences

$$c_k^2 = (2^{k+1} - 1)c_{k+1} \tag{3.2}$$

$$d_k c_k = (2^k - 1)d_{k+1} \tag{3.3}$$

$$2e_k c_k + d_k^2 = (2^{k+1} - 3)e_{k+1}. \tag{3.4}$$

We solve (3.2) subject to $c_1 = 1$. Setting $c_k = e^{u_k}$ leads to

$$u_{k+1} - 2u_k = -\log(2^{k+1} - 1) = -(k+1)\log 2 - \log(1 - 2^{-k-1}).$$

Solving this linear difference equation yields

$$u_k = (k+2)\log 2 + K2^k - \sum_{i=2}^{k} 2^{k-i} \log(1 - 2^{-i}).$$

where K is a constant. Now, $c_1 = 1$ implies that $u_1 = 0$ and hence $K = -\frac{3}{2}\log 2$. we thus have

$$c_k = L_{2^k-1}^k = 2^{k+2-3\cdot 2^{k-1}} \prod_{i=2}^{k} (1 - 2^{-i})^{-2^{k-i}}. \tag{3.5}$$

From (3.5) it follows immediately that as $k \to \infty$

$$\begin{aligned} c_k &= 4 \cdot 2^k \exp\left[-2^k \left(\log K_0 + \frac{3}{2}\log 2\right)\right] \left(1 + O(2^{-k})\right) \\ &\sim 4 \cdot 2^k \left(\frac{1}{2\sqrt{2}K_0}\right)^{2^k} = 4 \cdot 2^k e^{-2^k c_*} \end{aligned} \tag{3.6}$$

where $c_* = \frac{3}{2}\log 2 + \log K_0$ and

$$K_0 = \prod_{i=2}^{\infty} (1 - 2^{-i})^{2^{-i}} = .9103\dots\ . \tag{3.7}$$

Once we know c_k we can easily solve the linear equation (3.3) to obtain

$$d_k = (2^k - 1)c_k. \tag{3.8}$$

We thus have

$$d_k = L^k_{2^k-2} \sim 4^{k+1} \left(\frac{1}{2\sqrt{2}K_0} \right)^{2^k} = 4 \cdot 4^k e^{-2^k c_*}, \qquad k \to \infty. \tag{3.9}$$

In a similar manner we can solve the linear recurrence (3.4) for e_k.

Now consider general (fixed) M and n, $k \to \infty$. While we could infer the asymptotic behavior of L^k_n by continuing to solve the sequence of equations for the coefficients in (3.1), we shall instead use the recurrence (2.2). We change variables from (k, n) to (M, n) with

$$L^k_n = W(M; n) = W(2^k - n; n). \tag{3.10}$$

Replacing k by $k - 1$ in (2.2) and noting that

$$L^k_{n+1} = W(M - 1; n + 1), \qquad L^{k-1}_\ell = W\left(\frac{M}{2} + \frac{n}{2} - \ell; \ell \right)$$

leads to

$$\begin{aligned} W(M-1; n+1) &= \frac{1}{n+1} \sum_{\ell=0}^{n} W\left(\frac{M}{2} + \frac{n}{2} - \ell; \ell \right) W\left(\frac{M}{2} - \frac{n}{2} + \ell; n - \ell \right) \\ &= \frac{1}{n+1} \sum_{j=-n/2}^{n/2} W\left(\frac{M}{2} - j; \frac{n}{2} + j \right) W\left(\frac{M}{2} + j; \frac{n}{2} - j \right) \\ &= \frac{1}{n+1} \sum_{j=-M/2}^{M/2} W\left(\frac{M}{2} - j; \frac{n}{2} + j \right) W\left(\frac{M}{2} + j; \frac{n}{2} - j \right). \end{aligned} \tag{3.11}$$

To obtain the last equality we have used the fact that $W(M; n) = 0$ for $M < 0$ (i.e., $L^k_n = 0$ for $n > 2^k$) to truncate the limits on the sum. It follows that for a fixed M, the number of non-zero terms in the sum in (2.2) is $O(1)$ as $n \to \infty$.

We already know that $W(0; n) = 0$ and this can also be concluded by setting $M = 0$ in (3.11). By setting $M = 2$ we obtain

$$W(1; n + 1) = \frac{1}{n+1} W^2\left(1; \frac{n}{2}\right). \tag{3.12}$$

This admits an asymptotic solution as $n \to \infty$ in the form

$$W(1;n) = c_k \sim 4ne^{-c_* n - c_*} \tag{3.13}$$

where c_* is an arbitrary constant. Setting $M = 3$ in (3.11) yields

$$W(2;n+1) = \frac{2}{n+1} W\left(2; \frac{n-1}{2}\right) W\left(1; \frac{n+1}{2}\right). \tag{3.14}$$

In view of (3.13), (3.14) admits an asymptotic solution

$$W(2;n) = d_k \sim n^2 e^{-c_* n} 4c' e^{-c_*}, \qquad n \to \infty \tag{3.15}$$

where c' is also arbitrary. Let us assume that for a fixed M we have

$$W(M;n) \sim n^M e^{-c_* n} f(M), \qquad n \to \infty. \tag{3.16}$$

Using (3.16) in (3.11) and simplifying the result for $n \to \infty$, we find that $f(M)$ satisfies the recurrence

$$e^{-c_*} f(M-1) = 2^{-M} \sum_{\ell=1}^{M-1} f(\ell) f(M-\ell), \qquad f(0) = 0. \tag{3.17}$$

The most general solution to (3.17) is

$$f(M) = 4e^{-c_*} \frac{1}{(M-1)!} (c')^{M-1}, \qquad M \geq 1. \tag{3.18}$$

We have thus obtained, for fixed M,

$$L_n^k \sim \frac{4e^{-c_*}}{(M-1)!} (c')^{M-1} n^M e^{-c_* n}, \qquad n \to \infty. \tag{3.19}$$

It remains only to determine the constants c_* and c'. By comparing (3.19) with $M = 1$ to (3.6) we see that

$$c_* = \log(2\sqrt{2}K_0).$$

Note that now $n = 2^k - 1 \sim 2^k$. Setting $M = 2$ and comparing (3.9) and (3.19) we conclude that

$$c' = e^{-c_*},$$

where we used $n = 2^k - 2 \sim 2^k$. The expression (3.19) is asymptotically equivalent to

$$L_{2^k - M}^k \sim \frac{4}{(M-1)!} 2^{kM} \left(\frac{1}{2\sqrt{2}K_0}\right)^{2^k}, \qquad k \to \infty. \tag{3.20}$$

The form (3.20) is somewhat more numerically accurate than (3.19). We also note that if we use $d_k = (2^k - 1)c_k$ along with (3.2) in (3.4) and let $e_k = c_k \tilde{e}_k$, we find that

$$\frac{2^{k+1}-3}{2^{k+1}-1}\tilde{e}_{k+1} = (2^k-1)^2 + 2\tilde{e}_k. \tag{3.21}$$

With $M = 3$, (3.20) yields

$$L^k_{2^k-3} \sim 2 \cdot 2^{3k} e^{-2^k c_*} \sim 2^{2k-1} c_k$$

so that $\tilde{e}_k \sim 2^{2k-1}$ as $k \to \infty$, and this also follows from (3.21).

For purposes of asymptotic matching we will need the behavior of (3.19) as $M \to \infty$. Using Stirling's formula this yields

$$L^k_n \sim 4\sqrt{\frac{M}{2\pi}} e^{-c_*(n+M)} \exp\left(M \log\left(\frac{n}{M}\right) + M\right) \tag{3.22}$$

for $M \to \infty$.

4 Left Tail

We consider L^k_n in the range where $n, k \to \infty$ in such a way that $\omega \equiv n2^{-k}$ is fixed, with $0 < \omega < 1$. Note that as $\omega \to 1^-$ we have $n = 2^k(1 + o(1))$ so we are moving into the region where the M-scale analysis applies. We also have $k - \log_2 n = -\log_2 \omega > 0$ and $O(1)$ in this range. We define

$$V(\omega; n) = V(n2^{-k}; n) = L^k_n \tag{4.1}$$

and note that

$$L^k_{n+1} = V\left(\omega + \frac{\omega}{n}; n+1\right), \qquad L^{k-1}_\ell = V\left(\omega\frac{2\ell}{n}; \ell\right). \tag{4.2}$$

Replacing k by $k-1$ in (2.2) and using (4.1) and (4.2) leads to

$$V\left(\omega + \frac{\omega}{n}; n+1\right) = \frac{1}{n+1}\sum_{\ell=0}^{n} V\left(\omega\frac{2\ell}{n}; \ell\right) V\left(\omega\frac{2(n-\ell)}{n}; n-\ell\right). \tag{4.3}$$

The initial condition (2.3) does not apply on this scale, since k is assumed large.

For fixed ω we analyze (4.3) by a WKB-type expansion [4, 21, 26]. As discussed above, in the WKB approximation one assumes a particular form of the solution with some unknown parameters and/or functions. After substituting the solution into the original equation one tries to determine the unknown parameters. In our case, we assume an asymptotic expansion of the form

$$V(\omega; n) = n^\gamma e^{-n\phi(\omega)}\left[A(\omega) + \frac{1}{n}A^{(1)}(\omega) + \frac{1}{n^2}A^{(2)}(\omega) + \dots\right] \tag{4.4}$$

with γ a constant. We comment that the ω-scale was also important in the analysis of digital trees, such as tries, b-tries, PATRICIA tries [15] and digital search trees [16]. For tries and b-tries we can obtain an exact expression for the corresponding probability distribution of the height. Evaluating this for $n \to \infty$ and ω fixed by the saddle point method yields an asymptotic series in the form (4.4) (with $\gamma = 0$). For the PATRICIA and DST models, which have not been solved exactly, we used an "ansatz" similar to (4.4) (cf. [15, 16]).

Using (4.4) we obtain

$$\begin{aligned} V\left(\omega + \frac{\omega}{n}; n+1\right) &\sim (n+1)^\gamma A\left(\omega + \frac{\omega}{n}\right) e^{-(n+1)\phi(\omega+\omega/n)} \\ &\sim n^\gamma A(\omega) e^{-n\phi(\omega)} e^{-\phi(\omega)} e^{-\omega\phi'(\omega)}. \end{aligned} \tag{4.5}$$

With (4.4) the sum in (4.3) becomes, for $n \to \infty$,

$$\begin{aligned} & n^{2\gamma} \sum_{\ell=0}^{n} \frac{1}{n} A\left(2\omega\frac{\ell}{n}\right) A\left(2\omega\left(1-\frac{\ell}{n}\right)\right) \left(\frac{\ell}{n}\right)^\gamma \left(1-\frac{\ell}{n}\right)^\gamma \\ & \times \exp\left(-n\left[\frac{\ell}{n}\phi\left(2\omega\frac{\ell}{n}\right) + \left(1-\frac{\ell}{n}\right)\phi\left(2\omega\left(1-\frac{\ell}{n}\right)\right)\right]\right) \\ & \sim n^{2\gamma} \int_0^1 x^\gamma (1-x)^\gamma A(2\omega x) A(2\omega(1-x)) e^{-n[x\phi(2\omega x)+(1-x)\phi(2\omega(1-x))]} dx, \end{aligned} \tag{4.6}$$

where we have used the Euler-MacLaurin formula to approximate the sum by an integral. By symmetry, the major contribution to the integral will come from $x \approx \frac{1}{2}$. We thus define

$$\Phi(x) = \Phi(x;\omega) := x\phi(2\omega x) + (1-x)\phi(2\omega(1-x))$$

and note that

$$\Phi\left(\frac{1}{2}\right) = \phi(\omega), \quad \Phi'\left(\frac{1}{2}\right) = 0, \quad \Phi''\left(\frac{1}{2}\right) = 8\omega\phi'(\omega) + 4\omega^2\phi''(\omega).$$

Then expanding the integrand in (4.6) about $x = \frac{1}{2}$ and using the Laplace method yields

$$n^{2\gamma}[A(\omega)]^2 e^{-n\phi(\omega)} \frac{4^{-\gamma}}{\sqrt{n}} \sqrt{\frac{2\pi}{8\omega\phi'(\omega) + 4\omega^2\phi''(\omega)}} \left(1 + O(n^{-1})\right). \tag{4.7}$$

Upon comparing (4.5) to (4.7) we see that the factors $e^{-n\phi(\omega)}$ cancel and then

$$\gamma = \frac{1}{2}, \tag{4.8}$$

$$A(\omega) = 2\sqrt{\frac{2\omega}{\pi}} \sqrt{\omega\phi''(\omega) + 2\phi'(\omega)} e^{-(\phi(\omega)+\omega\phi'(\omega))}. \tag{4.9}$$

We have thus determined the algebraic factor n^γ in (4.4) and expressed $A(\omega)$ in terms of $\phi(\omega)$. However, we have not determined the latter function, which gives the exponential decay rate (in n) of the distribution on the ω-scale. The function $\phi(\omega)$ seems to be very sensitive to the initial conditions in (2.3), and we believe that it is unlikely that it can be determined solely from the recursion (2.2). By obtaining higher order terms in the approximations in (4.5) and (4.6) (using (4.4)) and then evaluating the higher order terms in the Laplace expansion of the integral, we can express $A^{(j)}(\omega)$ in terms of $\phi(\omega)$, $A(\omega)$, $A^{(1)}(\omega), \ldots,$ and $A^{(j-1)}(\omega)$. We can thus obtain the full asymptotic series, up to the function $\phi(\omega)$. This must be determined numerically, and this we do in Section 9.

We can obtain analytically the behavior of $\phi(\omega)$ as $\omega \to 0$ and $\omega \to 1$, by using the asymptotic matching principle. The limit $\omega \to 0^+$ will be discussed in Section 8, where we analyze the "central region". Now consider the intermediate limit where $\omega = n2^{-k} \to 1^-$, but $M = 2^k - n \to \infty$. Noting that $\frac{n}{M} = \frac{\omega}{1-\omega} = \frac{1}{1-\omega} - 1$ we rewrite (3.22) in terms of n and ω, which yields

$$\begin{aligned} L_n^k \quad &\sim \quad 4\sqrt{\frac{n(1-\omega)}{2\pi}} \exp\left[-nc_* - n(1-\omega)\log(1-\omega) - n(1-\omega)(c_* - 1)\right. \\ &+ \quad \left. O(n(1-\omega)^2 \log(1-\omega))\right] (1 + O(1-\omega))\,. \end{aligned} \tag{4.10}$$

Choosing an intermediate limit where $\omega \to 1$ and $n \to \infty$ with $n(1-\omega)^2\log(1-\omega) \to 0$ (e.g., $1 - \omega = O(n^{-2/3})$), the matching condition implies that as $\omega \to 1$, $\sqrt{n}A(\omega)e^{-n\phi(\omega)}$ should agree with (4.10), which implies that

$$\phi(\omega) = c_* + (1-\omega)\log(1-\omega) + (1-\omega)(c_* - 1) + o(1-\omega) \tag{4.11}$$

and

$$A(\omega) \sim 4\sqrt{\frac{1-\omega}{2\pi}}, \qquad \omega \to 1. \tag{4.12}$$

This shows that $\phi(\omega)$ is finite at $\omega = 1$, with $\phi(1) = c_* = \log(2\sqrt{2}K_0) = .9457\ldots$, but its derivative has a logarithmic singularity at $\omega = 1$. The asymptotic matching condition yielded independently the behavior of $\phi(\omega)$ and $A(\omega)$ as $\omega \to 1$. Note also that the matching is only possible if $\gamma = \frac{1}{2}$ in (4.4). We show that (4.11) and (4.12) are indeed consistent with the relationship between $\phi(\omega)$ and $A(\omega)$ in (4.8). By differentiating (4.11) we obtain

$$\phi'(\omega) \sim -\log(1-\omega) - c_*, \quad \phi''(\omega) \sim \frac{1}{1-\omega}, \quad \omega \to 1$$

and hence as $\omega \to 1$

$$\begin{aligned} 2\sqrt{\frac{2\omega}{\pi}}\sqrt{\omega\phi''(\omega) + 2\phi'(\omega)} \quad &\sim \quad \frac{2\sqrt{2}}{\sqrt{\pi(1-\omega)}}, \\ e^{-\phi(\omega)}e^{-\omega\phi'(\omega)} \quad &\sim \quad e^{-c_*}e^{\log(1-\omega)+c_*} = 1 - \omega. \end{aligned}$$

Thus (4.8) agrees with (4.12).

Finally we comment that for the digital trees we studied in [15, 16], the corresponding expansion (4.4) had $\gamma = 0$ and the corresponding $A(\omega)$ satisfied $A(\omega) \to 1$ as $\omega \to 0$. Note that as $\omega \to 0$ we have $k - \log_2 n \to \infty$ so we are approaching the region where the mass is concentrated, which occurs for $k = A(\log n)[1 + o(1)]$ and $A > 1/\log 2$. In the present case we have $\gamma = \frac{1}{2}$ and we will show in Section 8 that $A(\omega)$ cannot $\to 1$ as $\omega \to 0$. Thus the structure of the binary search tree model is much different in the central region, than digital trees such as PATRICIA and DST.

5 Far Right Tail

We proceed to analyze the right tail regions, starting with the far right tail where $n - k = j = O(1)$. We shall use *linearization* of (2.2) to obtain asymptotic results in this region. Our strategy is to move toward the central region from both the left and right sides, so we analyze the central region last.

We change variables from (k, n) to (j, n) with $j = n - k$ and

$$L_n^k = L_n^{n-j} = 1 - F_n^j. \tag{5.1}$$

From (2.2) we then obtain the following problem for F_n^j:

$$\begin{aligned} F_{n+1}^j &= \frac{2}{n+1}\sum_{\ell=0}^{n} F_\ell^{\ell+j-n} - \frac{1}{n+1}\sum_{\ell=0}^{n} F_\ell^{\ell+j-n} F_{n-\ell}^{j-\ell} \\ &= \frac{2}{n+1}\sum_{\ell=n-j+1}^{n} F_\ell^{\ell+j-n} - \frac{1}{n+1}\sum_{\ell=n-j+1}^{j-1} F_\ell^{\ell+j-n} F_{n-\ell}^{j-\ell}. \end{aligned} \tag{5.2}$$

Here we have used the fact that $F_n^j = 0$ for $j \leq 0$ (since $L_n^k = 1$ for $k \geq n$) to truncate the limits on the sums in (5.2). If $2j < n + 2$ then the second sum becomes void and the equation becomes linear:

$$F_{n+1}^j = \frac{2}{n+1}\sum_{\ell=0}^{j-1} F_{n-\ell}^{j-\ell}, \qquad j < \frac{n}{2} + 1. \tag{5.3}$$

We can easily solve (5.2) for small values of j. Setting $j = 1$ we have

$$F_{n+1}^1 = \frac{2}{n+1} F_n^1, \qquad n \geq 1. \tag{5.4}$$

Since $L_1^0 = 0$ we have $F_1^1 = 1$ and hence the solution to (5.4) is

$$F_n^1 = \frac{2^{n-1}}{n!}. \tag{5.5}$$

For $j = 2$ (5.3) then yields

$$F_{n+1}^2 = \frac{2}{n+1}[F_n^2 + F_{n-1}^1] = \frac{2}{n+1}F_n^2 + \frac{2^{n-1}}{(n+1)}\frac{1}{(n-1)!}, \qquad n \geq 3. \tag{5.6}$$

Solving this simple difference equation gives

$$F_n^2 = \frac{2^n}{n!}\left[\frac{n(n-1)}{8} + c_2\right], \qquad n \geq 3. \tag{5.7}$$

To determine c_2 we use (5.2) with $j = 2$ and $n = 1, 2, 3$ to conclude that $F_2^2 = F_3^2 = F_4^2 = 1$. Using (5.7) with $n = 3$ yields $c_2 = 0$ and hence

$$1 - L_n^{n-2} = \frac{2^{n-3}}{(n-2)!}, \qquad n \geq 3. \tag{5.8}$$

We set $j = 3$ and use (5.3) for $n \geq 5$ so that

$$F_{n+1}^3 = \frac{2}{n+1}[F_n^3 + F_{n-1}^2 + F_{n-2}^1], \qquad n \geq 5. \tag{5.9}$$

In view of (5.5) and (5.7) we have

$$F_{n+1}^3 - \frac{2}{n+1}F_n^3 = \frac{2}{n+1}\left[\frac{2^{n-4}}{(n-3)!} + \frac{2^{n-3}}{(n-2)!}\right] \tag{5.10}$$

whose solution is

$$F_n^3 = \frac{2^n}{n!}\left[\frac{n^4 + 3n^2}{64} - \frac{5n^3 + n}{96} + c_3\right], \qquad n \geq 5. \tag{5.11}$$

By examining (5.2) with $j = 3$ and $n = 2, 3, 4, 5$ we find that $F_3^3 = F_4^3 = F_5^3 = 1$. It follows that $c_3 = -5/8$ and hence

$$F_n^3 = \frac{2^{n-6}}{n!}\left[n(n-1)(n-2)\left(n - \frac{1}{3}\right) - 40\right], \qquad n \geq 5. \tag{5.12}$$

In a similar manner we can compute F_n^j for any fixed j.

For a fixed j and n sufficiently large F_n^j has the form $2^n(n!)^{-1}\times$ [polynomial in n of degree $2j-2$]. We thus write

$$F_n^j \sim \frac{2^n}{n!}n^{2j-2}\alpha_j, \qquad n \to \infty. \tag{5.13}$$

Using (5.13) we see that only the terms with $\ell = 0$ and $\ell = 1$ in the sum in (5.3) are asymptotically important. We thus write (5.3) as

$$F_{n+1}^j \sim \frac{2}{n+1}[F_n^j + F_{n-1}^{j-1}]$$

and use (5.13), which yields the recurrence

$$(2j-2)\alpha_j = \frac{1}{2}\alpha_{j-1}, \qquad \alpha_1 = \frac{1}{2} \tag{5.14}$$

so that $\alpha_j = \frac{2^{1-2j}}{(j-1)!}$ and

$$1 - L_n^{n-j} \sim \frac{2^{n-1}}{n!}\left(\frac{n}{2}\right)^{2j-2}\frac{1}{(j-1)!}.$$

We have thus obtained the leading term in the far right tail region. Note that the error term in (5.13) is $1 + O(n^{-1})$.

Our analysis also shows that the non-linear equation becomes exactly linear for $n > 2(j-1)$. This corresponds (roughly) to the sector $\frac{n}{2} < k < n$ of the (n,k) plane. We thus refer to this as the "linear sector", and to the range $0 < k < \frac{n}{2}$ as the "non-linear sector". To uniquely determine the solution to (5.3) in the linear sector we need the values of F_n^j when $n = 2j - 1$. Unfortunately, these ultimately depend on how the initial conditions in (2.3) propagate through the non-linear sector. Thus (5.2) and (5.3) are not immediately useful for determining F_n^j *exactly*. However, throughout the right tail of the distribution we have $L_n^k \sim 1$ and thus F_n^j is asymptotically small. Then the non-linear term in (5.2) is asymptotically small compared to the linear part and we write

$$F_{n+1}^j \sim \frac{2}{n+1}\sum_{\ell=0}^{j-1} F_{n-\ell}^{j-\ell} \quad \text{for} \quad F_n^j << 1 \;\; \text{as} \;\; n \to \infty. \tag{5.15}$$

In Section 8 we show that the non-linear part becomes negligible as we move out of the central region and into the right tail.

Finally, we note that if we expand the leading term for F_n^j for j fixed as $j \to \infty$, we obtain

$$F_n^j \sim n^{-n} j^{-j} 2^n e^{n+j} \frac{\sqrt{j}}{\pi n^{5/2}} n^{2j} 2^{-2j}, \quad j \to \infty. \tag{5.16}$$

This will be used for asymptotic matching purposes, as in the next section we will show that (5.13) ceases to be valid if $n, j \to \infty$ at the same rate.

6 Right Tail

We consider the limit $n, k \to \infty$ with k/n fixed and $0 < k/n < 1$. We define $\alpha = n/j = n/(n-k)$ so that α is fixed and > 1. First, we make some observations about the *general* solution to the linear problem (5.15). Upon setting

$$F_n^j = \frac{2^n}{n!} G_n^j \tag{6.1}$$

we obtain

$$G_{n+1}^{j} = \sum_{\ell=0}^{j-1} \frac{n!}{(n-\ell)!} 2^{-\ell} G_{n-\ell}^{j-\ell}. \tag{6.2}$$

Setting $j = 1$ yields $G_n^1 = \bar{c}_1$, a constant. Then setting $j = 2, 3, \ldots$ and recursively solving the resulting difference equations yields

$$\begin{aligned}
G_n^2 &= \frac{1}{4}\bar{c}_1 n(n-1) + c_2, &(6.3)\\
G_n^3 &= \frac{1}{32}\bar{c}_1 n(n-1)(n-2)(n-3) + \frac{1}{12}\bar{c}_1 n(n-1)(n-2) + \frac{1}{4}c_2 n(n-1) + c_3, &(6.4)\\
G_n^4 &= \frac{1}{384}\bar{c}_1 n(n-1)(n-2)(n-3)(n-4)(n-5) + \frac{5}{240}\bar{c}_1 n(n-1)(n-2)(n-3)(n-4) &(6.5)\\
&+ \frac{1}{32}\bar{c}_1 n(n-1)(n-2)(n-3) + \frac{1}{32}c_2 n(n-1)(n-2)(n-3)\\
&+ \frac{1}{12}c_2 n(n-1)(n-2) + \frac{1}{4}c_3 n(n-1) + c_4,
\end{aligned}$$

and so forth. It is clear that for general j the solution has the form

$$G_n^j = \bar{c}_1 P_{2j-2}(n) + c_2 Q_{2j-4}(n) + c_3 R_{2j-6}(n) + \cdots \tag{6.6}$$

where P, Q and R and polynomials in n of respective degrees $2j-2$, $2j-4$ and $2j-6$. We write these polynomials as

$$\begin{aligned}
P_{2j-2}(n) &= \sum_{m=0}^{j-2} A_m^j [n(n-1)(n-2)\cdots(n-2j+m+3)],\\
Q_{2j-4}(n) &= \sum_{m=0}^{j-3} B_m^j [n(n-1)(n-2)\cdots(n-2j+m+5)],\\
R_{2j-6}(n) &= \sum_{m=0}^{j-4} C_m^j [n(n-1)(n-2)\cdots(n-2j+m+7)]. &(6.7)
\end{aligned}$$

Furthermore, the coefficients are related by

$$\begin{aligned}
B_m^j &= A_m^{j-1}, & j \geq 3,\\
C_m^j &= B_m^{j-1} = A_m^{j-2} & j \geq 4.
\end{aligned}$$

In Appendix A we estimate the relative sizes of these polynomials in various limits. We consider

$$\frac{Q_{2j-4}(n)}{P_{2j-2}(n)}, \quad \frac{R_{2j-6}(n)}{P_{2j-2}(n)}$$

and show that for $n \to \infty$ and j fixed we have $Q/P = O(n^{-2})$ and $R/P = O(n^{-4})$. For $n, j \to \infty$ with $\alpha = n/j$ fixed and > 1, we obtain $Q/P = O(n^{-1})$ and $R/P = O(n^{-2})$.

However, if $k,n \to \infty$ with $\nu = k/(\log n)$ fixed and $\nu > A = 4.311\ldots$, then P, Q and R become of comparable magnitude.

Let us define

$$\tilde{F}_n^j = \bar{c}_1 \frac{2^n}{n!} P_{2j-2}(n). \tag{6.8}$$

From our results in Section 5, we see that on the j scale we have $F_n^j \sim \tilde{F}_n^j$ and we refer to $\tilde{F}_n^j$ as the "uniform right tail" (URT) approximation. It applies in the right tail with the exception of the near right tail, where all the terms in the series (6.6) contribute. Thus we need $k/\log n \to \infty$ for the URT result to hold. We shall analyze this case separately in Section 7. The calculations in Section 5 showed that $\bar{c}_1 = \frac{1}{2}$, $c_2 = 0$ and $c_3 = -\frac{5}{8}$. To obtain these values we needed to use the initial conditions in (2.3) and to see how they propagate through the non-linear sector. While this method can compute the first few c_ℓ, it doesn't seem feasible to obtain c_ℓ in general. Fortunately, only $\bar{c}_1$ is important in most of the right tail region. Since $c_2 = 0$ the above discussion shows that $F_j^n/\tilde{F}_j^n = 1 + O(n^{-4})$ if $n \to \infty$ with j fixed, and $F_j^n/\tilde{F}_j^n = 1 + O(n^{-2})$ on the α-scale. We proceed to explicitly calculate the polynomial P.

We use (6.6) and (6.7) in (6.2) to obtain (neglecting c_ℓ for $\ell \geq 2$)

$$\sum_{m=0}^{j-2} A_m^j (n+1)n(n-1)\cdots(n+m-2j+4) - \sum_{m=0}^{j-2} A_m^j n(n-1)(n-2)\cdots(n+m-2j+3)$$
$$= \sum_{\ell=1}^{j-1}\sum_{m=0}^{j-\ell-2} A_m^{j-\ell} 2^{-\ell} n(n-1)(n-2)\cdots(n+m+\ell-2j+3).$$

By comparing coefficients of $(n)_\ell = n(n-1)\cdots(n-\ell+1)$ we are led to the recurrence

$$(2j-m)A_{m-2}^j = \frac{2^{1-m}}{(j-m)!}4^{m-j} + \sum_{\ell=1}^{m-2} 2^{-\ell} A_{m-\ell-1}^{j-\ell}, \qquad j \geq m+1. \tag{6.9}$$

Note that A_m^j is defined for $0 \leq m \leq j-2$, and we have used $A_0^{j-m+1} = 4^{m-j}/(j-m)!$ to isolate the term with $\ell = m-1$ in the above sum. From (6.9) we obtain the boundary values

$$\begin{aligned} A_{m-2}^m &= \frac{2^{1-m}}{m}, \quad m \geq 2, \\ A_0^j &= \frac{4^{1-j}}{(j-1)!} \quad j \geq 2. \end{aligned}$$

We also note that as $n \to \infty$ with j fixed $G_n^j \sim \bar{c}_1 A_0^j n^{2j-2}$, which when used in (6.1) regains the leading term in the far right tail.

To solve (6.9) we set $\ell = j - m - 2$ with

$$A_m^j = B(m, j-m-2) = B(m,\ell) \tag{6.10}$$

to obtain (shifting $m \to m+2$ in (6.9))

$$(m+2\ell+2)B(m,\ell) = 2^{-m-1}\frac{4^{-\ell}}{\ell!} + \sum_{p=1}^{m} 2^{-p}B(m-p+1,\ell-1) \tag{6.11}$$

and this holds for $m, \ell \geq 0$. Furthermore we let

$$B(m,\ell) = 4^{-\ell}2^{-m-1}H(m,\ell)$$

and obtain from (6.11)

$$[2(\ell+1)+m]H(m,\ell) = \frac{1}{\ell!} + 2\sum_{p=1}^{m} H(m-p+1,\ell-1); \quad m,\ell \geq 0. \tag{6.12}$$

We next introduce the bivariate generating function

$$\tilde{H}(z,w) = \sum_{m=0}^{\infty}\sum_{\ell=0}^{\infty} z^m w^\ell H(m,\ell) \tag{6.13}$$

and obtain from (6.12) the partial differential equation

$$z\frac{\partial \tilde{H}(z,w)}{\partial z} + 2\frac{\partial}{\partial w}(w\tilde{H}(z,w)) = \frac{e^w}{1-z} + \frac{2w}{1-z}(\tilde{H}(z,w) - \tilde{H}(0,w)).$$

Letting $z \to 0$ yields

$$2\frac{\partial}{\partial w}(w\tilde{H}(0,w)) = e^w$$

which implies that

$$\tilde{H}(0,w) = \frac{e^w - 1}{2w},$$

and then the above partial differential equation becomes

$$z\frac{\partial \tilde{H}(z,w)}{\partial z} + 2\frac{\partial}{\partial w}(w\tilde{H}(z,w)) = \frac{1}{1-z} + \frac{2w}{1-z}\tilde{H}(z,w). \tag{6.14}$$

Letting $w \to 0$ yields

$$\tilde{H}(z,0) = -\frac{1}{z^2}[z + \log(1-z)] = \sum_{m=0}^{\infty} \frac{z^m}{m+2}.$$

Using the sequence of substitutions $z = e^{-s}$, $w = e^{-t}$, and then $a = 2s - t$, $b = s$, with $\tilde{H}(z,w) = \bar{H}(a,b)$, we arrive at the following simpler PDE

$$\frac{\partial \bar{H}(a,b)}{\partial b} + \bar{H}(a,b)\left(\frac{2e^{a-2b}}{1-e^{-b}} - 2\right) = \frac{-1}{1-e^{-b}}$$

that can be solved leading to the following solution analytic at $z = 0$, $w = 0$

$$\tilde{H}(z,w) = \frac{1}{2w}\left[\exp\left(-\frac{2w}{z^2}[z+\log(1-z)]\right) - 1\right]. \tag{6.15}$$

Inverting the transform over w in (6.13) we find that

$$\sum_{m=0}^{\infty} z^m H(m,\ell) = \left(\frac{-\log(1-z)-z}{z^2}\right)^{\ell+1} \frac{2^\ell}{(\ell+1)!}.$$

By Cauchy's formula (cf. [26]) and (6.10) we obtain

$$A_m^j = \frac{2^{1-j}}{(j-m-1)!}\frac{1}{2\pi i}\oint z^{-m-1}\left(\frac{-\log(1-z)-z}{z^2}\right)^{j-m-1} dz \tag{6.16}$$

where the loop integral is around $z = 0$. Using $\bar{c}_1 = \frac{1}{2}$ we thus have, from (6.7), (6.9) and (6.16),

$$\tilde{F}_n^j = 2^{n-j}\sum_{m=0}^{j-2}\frac{1}{(j-m-1)!}\frac{1}{(n-2j+m+2)!}I(m,j) \tag{6.17}$$

where

$$I(m,j) = \frac{1}{2\pi i}\oint z^{-m-1}\left(\frac{-\log(1-z)-z}{z^2}\right)^{j-m-1} dz. \tag{6.18}$$

We proceed to evaluate $\tilde{F}_n^j$ asymptotically in various limits. Since $j = n-k$ we see that $\tilde{F}_n^j$ is defined for all $k \geq 0$. For $n < 2j-2$ (i.e., $n > 2(k+1)$), the lower limit on the sum in (6.17) must be truncated at $m = 2j-2-n$. We first establish the following lemma.

Lemma 1 *The Cauchy integral $I(m,j)$ has the following asymptotic expansions.*
(1) *For $m \to \infty$, $j - m$ fixed*

$$I \sim \frac{j-m-1}{m}(\log m)^{j-m-2}.$$

(2) *For $m \to \infty$, $j - m \to \infty$ with $(j-m)/\log m = \beta_*$, $0 < \beta_* < \infty$*

$$I \sim \frac{1}{m}(\log m)^{j-m-1}\frac{e^{-\beta_*}}{\Gamma(\beta_*)}$$

where $\Gamma(\cdot)$ is the Euler gamma function.
(3) *For $m, j \to \infty$ with m/j fixed and $0 < m/j < 1$*

$$I \sim \frac{1}{\sqrt{2\pi}} z_0^{-m-1}[\Delta(z_0)]^{j-m-1}\frac{1}{\sqrt{2j-m}}\left[\frac{1}{(1-z_0)z_0} - \frac{m}{j-m}\frac{1}{z_0^2}\right]^{-1/2},$$

where $z_0 = z_0(m/j)$ is the solution to

$$1 + \frac{j}{j-m} = \frac{z_0^2}{(1-z_0)}\frac{-1}{\log(1-z_0)+z_0} = \frac{1}{1-z_0}\frac{1}{\Delta(z_0)}$$

and $\Delta(z) = -[\log(1-z) + z]/z^2$.

(4) *For* $j \to \infty$, m *fixed*

$$I \sim 2^{m+1-j} \frac{1}{m!} \left(\frac{2}{3} j\right)^m .$$

Proof. Since the result follows from asymptotically evaluating the integral in (6.18) using standard methods, we only briefly sketch the proof. We note that other than $z = 0$ the integrand's only singularity is the branch point at $z = 1$.

For part (1) we use the singularity analysis of Flajolet and Odlyzko [11]). Observe that

$$I(m,j) = [z^m] \left(\frac{-\log(1-z) - z}{z^2}\right)^{j-m-1}$$

where $[z^n]f(z)$ stands for the coefficient at z^n in the Laurent expansion of $f(z)$. Since $\beta = j - m - 1$ is fixed, the singularity analysis is applicable. We know that (cf. [11, 26])

$$[z^m] \left(\frac{-\log(1-z)}{z}\right)^\beta = \frac{(\log m)^\beta}{m} \left(\frac{\beta}{\log m} + O(1/\log^2 m)\right).$$

Thus

$$[z^m] \left(\frac{-\log(1-z) - z}{z}\right)^\beta \sim \frac{\beta (\log m)^{\beta-1}}{m},$$

and this implies that

$$[z^m] \left(\frac{-\log(1-z) - z}{z^2}\right)^\beta = [z^{m-\beta}] \left(\frac{-\log(1-z) - z}{z}\right)^\beta \sim \frac{\beta (\log m)^{\beta-1}}{m}$$

which proves part (1) of the lemma.

For $j \to \infty$ and m fixed the major contribution to the integral in (6.18) comes from $z = 0$ and we obtain

$$\begin{aligned} I &= \frac{1}{2\pi i} \oint z^{-m-1} \left(\frac{1}{2} + \frac{z}{3} + O(z^2)\right)^{j-m-1} dz \qquad (6.19) \\ &\sim 2^{m+1-j} \frac{1}{2\pi i} \oint z^{-m-1} e^{2zj/3} dz \\ &= 2^{m+1-j} \frac{1}{m!} \left(\frac{2}{3} j\right)^m \end{aligned}$$

and this establishes Lemma 1 part (4).

Now consider $m, j \to \infty$ at the same rate. We use (6.18) and the saddle point method (cf. [26, 27]), writing the integrand as $[z\Delta(z)]^{-1} \exp[-m \log z + (j-m)\log(\Delta(z))]$. The saddle point equation is

$$\frac{d}{dz} \left(-m \log z + (j-m) \log(\Delta(z))\right) = 0$$

or

$$\frac{m}{j-m}=z\frac{\Delta'(z)}{\Delta(z)}=-2+\frac{1}{1-z}\frac{1}{\Delta(z)}.$$

We can easily show that the above has a unique solution $z=z_0=z_0(m/j)$ that lies on the real axis in the range $0<z<1$. The directions of steepest descent at this saddle are $\arg(z-z_0)=\pm\pi/2$ and then the standard Laplace estimate yields

$$I\sim\frac{1}{z_0\Delta(z_0)}\frac{1}{\sqrt{2\pi}}z_0^{-m}[\Delta(z_0)]^{j-m}\left(\frac{m}{z_0^2}+(j-m)\left[\frac{\Delta''(z_0)}{\Delta(z_0)}-\left(\frac{\Delta'(z_0)}{\Delta(z_0)}\right)^2\right]\right)^{-1/2}. \tag{6.20}$$

Now, $z\Delta'(z)+2\Delta(z)=1/(1-z)$ whose derivative yields

$$\begin{aligned} z_0\frac{\Delta''(z_0)}{\Delta(z_0)} &= -3\frac{\Delta'(z_0)}{\Delta(z_0)}+\frac{1}{(1-z_0)^2}\frac{1}{\Delta(z_0)} \\ &= -3\frac{m}{j-m}\frac{1}{z_0}+\frac{1}{(1-z_0)}\frac{2j-m}{j-m}. \end{aligned}$$

It follows that

$$\frac{m}{z_0^2}+(j-m)\left[\frac{\Delta''(z_0)}{\Delta(z_0)}-\left(\frac{\Delta'(z_0)}{\Delta(z_0)}\right)^2\right]=(2j-m)\frac{1}{z_0}\left[\frac{1}{1-z_0}-\frac{m}{j-m}\frac{1}{z_0}\right].$$

Using the above in (6.20) establishes part (3) of the Lemma. We note that $z_0\to 0^+$ as $m/j\to 0$ and $z_0\to 1^-$ as $m/j\to 1$; more precisely

$$z_0\sim\frac{3}{2}\frac{m}{j},\ \frac{m}{j}\to 0;\quad 1-z_0\sim\frac{-\delta}{\log\delta},\delta=1-\frac{m}{j}\to 0.$$

Finally, we prove part (2). We set $j-m=\beta_*\log m$. By deforming the contour in (6.18) into an integral about the branch cut (i.e., Hankel contour; cf. [26, 27]), we obtain the following alternate representation for I:

$$I=\frac{1}{2\pi i}\int_{+\infty}^{(0^-)}(1+y)^{-m-1}\left[\frac{\pi i-\log y-1-y}{(1+y)^2}\right]^{j-m-1}dy \tag{6.21}$$

where $\int_{+\infty}^{(0^-)}$ denotes the Hankel integral along a path starting at infinity on the lower half-plane, winding clockwise around the origin and proceeding back to infinity. Using $y=u/m$ we thus obtain

$$\begin{aligned} I &\sim \frac{1}{2\pi i m}\int_{+\infty}^{(0^-)}e^{-u}(\log m)^{j-m-1}\left[1+\frac{\pi i-\log u-1}{\log m}+O(m^{-1})\right]^{\beta_*\log m-1}du \\ &\sim \frac{e^{-\beta_*}(\log m)^{j-m-1}}{m}\frac{1}{2\pi i}\int_{+\infty}^{(0^-)}e^{-u}(-u)^{-\beta_*}du \\ &= \frac{e^{-\beta_*}(\log m)^{j-m-1}}{m\Gamma(\beta_*)}, \end{aligned}$$

since (cf. [1, 2, 26])

$$\frac{1}{2\pi i}\int_{+\infty}^{(0^-)} e^{-u}(-u)^{-s}du = \frac{1}{\Gamma(s)}.$$

This proves the entire Lemma 1. ■

With Lemma 1 we have the expansion of $I(m,j)$ for all possible ranges in the sum in (6.17). We next obtain the asymptotic expansion of $\tilde{F}_n^j$ in various ranges; these are summarized below.

Lemma 2 *The asymptotic expansions of $\tilde{F}_n^j$ in (6.17) are as follows:*
(a) *For $n \to \infty$, $j = O(1)$*

$$\tilde{F}_n^j \sim \frac{2^n}{n!}\frac{n^{2j-2}}{(j-1)!}2^{1-2j} \sim \frac{1}{\sqrt{2\pi n}}\left(\frac{2e}{n}\right)^n \frac{n^{2j-2}}{(j-1)!}2^{1-2j}.$$

(b) *For $n, j \to \infty$, $j = O(\sqrt{n})$*

$$\begin{aligned}\tilde{F}_n^j &\sim \frac{1}{n!}\frac{1}{(j-1)!}2^{n+1-2j}n^{2j-2}\exp\left(-\frac{2}{3}\frac{j^2}{n}\right)\\ &\sim \frac{1}{\pi}\sqrt{\frac{j}{n}\frac{e^{n+j}}{n^n j^j}}2^{n-2j}n^{2j-2}\exp\left(-\frac{2}{3}\frac{j^2}{n}\right).\end{aligned}$$

(c) *For $n, j \to \infty$, $\alpha = n/j$ fixed, $1 < \alpha < \infty$*

$$\tilde{F}_n^j \sim \frac{(2e)^{n-j}}{2\pi n^2}n^{j-n}z_*^{-j}(1-z_*)^{j-n}\frac{\sqrt{\alpha}z_*}{\sqrt{1-z_*}\sqrt{\alpha z_*-1}},$$

where $z_* = z_*(\alpha)$ is the unique solution to

$$\frac{1}{\alpha} = \sum_{m=1}^{\infty}\frac{z_*^m}{m(m+1)} = 1 + \frac{1}{z_*}(1-z_*)\log(1-z_*), \qquad 0 < z_* < 1.$$

(d) *For $n, k \to \infty$, $k = O(\log n)$*

$$\tilde{F}_n^j \sim \left(\frac{2e}{k}\right)^k \frac{(\log n)^{k+1}}{nk^{3/2}\sqrt{2\pi}}\frac{1}{\Gamma\left(\frac{k}{\log n}\right)}.$$

(e) *For $k = O(1)$, $n \to \infty$*

$$\tilde{F}_n^j \sim \frac{2^k}{k!}\frac{(\log n)^k}{n}.$$

Proof. We again only sketch the proof of Lemma 2. If j is fixed and $n \to \infty$ the dominant term asymptotically in the sum in (6.17) corresponds to $m = 0$. Since $I(0,j) = 2^{1-j}$ we obtain part (a) of Lemma 2.

Now let $n, j \to \infty$ with $j/\sqrt{n}$ fixed. We write

$$\tilde{F}_n^j = \frac{2^{n-1}}{n!} \sum_{m=0}^{j-2} f(m)$$

where

$$f(m) = \frac{A_m^j}{(n-2j+m+2)!}.$$

Note that

$$\frac{f(m+1)}{f(m)} = \frac{A_{m+1}^j}{A_m^j} \frac{1}{n-2j+m+3}.$$

The main contribution of the sum comes from $m = O(1)$. By part (4) of Lemma 1 we conclude that for $m = O(1)$ and $j = O(\sqrt{n})$

$$\frac{f(m+1)}{f(m)} \sim \frac{4}{3} \frac{j^2}{nm}.$$

Now all terms with $m = O(1)$ in the sum in (6.17) are of comparable magnitude, and we obtain

$$\begin{aligned} \tilde{F}_n^j &\sim \frac{2^{n+1-2j}}{(j-1)!} \sum_{m=0}^{j-2} \left(\frac{4}{3} j^2\right)^m \frac{1}{m!} \frac{1}{n^{2+m}(n-2j)!} \\ &\sim \frac{2^{n+1-2j}}{(j-1)!} \frac{1}{n^2} \frac{1}{(n-2j)!} \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{4j^2}{3n}\right)^m. \end{aligned}$$

Summing the exponential series and using $(n-2j)! \sim n! n^{-2j} \exp(2j^2/n)$ leads to part (b) of Lemma 2.

Next we consider $n, j \to \infty$ at the same rate with $\alpha = n/j > 1$. Now the dominant contribution to the sum in (6.17) will come from large m, with $m = O(j)$. In this range we may use the expansion in part (3) of Lemma 1. Note also that in this limit

$$\frac{f(m+1)}{f(m)} \sim \frac{j-m}{n+m-2j} \frac{1}{z_0 \Delta(z_0)} = \frac{2j-m}{n+m-2j} \frac{1-z_0}{z_0}$$

where we have used the equation satisfied by $z_0(m/j)$. As $m/j \to 0$ we have $z_0 \to 0$ and $f(m+1) > f(m)$; while for $m/j \to 1$ we have $z_0 \to 1$ and $f(m+1) < f(m)$. Thus the major contribution to the sum should come from where $f(m+1)/f(m) = 1$, so that

$$z_0 = z_0\left(\frac{m}{j}\right) = \frac{2j-m}{n}. \tag{6.22}$$

We use Stirling's formula in (6.17) in the form

$$\begin{aligned} \frac{1}{(j-m-1)!} &\sim \sqrt{\frac{j-m}{2\pi}} e^{(m-j)\log(j-m)} e^{j-m}, \\ \frac{1}{(n+m-2j+2)!} &\sim \frac{1}{\sqrt{2\pi}} \frac{1}{(n+m-2j)^{5/2}} e^{(2j-n-m)\log(n+m-2j)} e^{n+m-2j} \end{aligned}$$

along with Lemma 1(3) to obtain

$$
\begin{aligned}
\tilde{F}_n^j &\sim \frac{2^{n-j}}{(2\pi)^{3/2}} \sum_{m=0}^{j-2} \sqrt{\frac{j-m}{2j-m}} \frac{1}{(n+m-2j)^{5/2}} \frac{1}{z_0\Delta(z_0)} \left[\frac{1}{(1-z_0)z_0} - \frac{m/j}{1-m/j} \frac{1}{z_0^2}\right]^{-1/2} \qquad (6.23) \\
&\times \exp\{n-j+(m-j)\log(j-m)+(2j-n-m)\log(n+m-2j) \\
&- m\log z_0 + (j-m)\log[\Delta(z_0)]\}.
\end{aligned}
$$

Next we set $\alpha = n/j > 1$, $x = m/j \in (0,1)$ and use the identity $\log[\Delta(z_0(x))] = \log(1-x) - \log(2-x) - \log[1-z_0(x)]$. Approximating (6.23) by an integral via Euler-MacLaurin, we arrive at

$$
\tilde{F}_n^j \sim \frac{2^{n-j}}{(2\pi)^{3/2}} e^{n-j} e^{(j-n)\log j} \frac{1}{j^{3/2}} \int_0^1 e^{jF(x;\alpha)} G(x;\alpha)\,dx \qquad (6.24)
$$

where

$$
\begin{aligned}
F(x;\alpha) &= (2-x-\alpha)\log(x+\alpha-2) - x\log[z_0(x)] - (1-x)\log[1-z_0(x)] + (x-1)\log(2-x) \\
G(x;\alpha) &= \frac{1-z_0(x)}{(\alpha+x-2)^{5/2} z_0(x)} \sqrt{\frac{2-x}{1-x}} \left[\frac{1}{z_0(x)} \frac{1}{1-z_0(x)} - \frac{x}{1-x} \frac{1}{z_0^2(x)}\right]^{-1/2}.
\end{aligned}
$$

By the Laplace method, the major contribution to the integral will come from where F is maximal in the range $0 \le x \le 1$. We have

$$
\frac{\partial F}{\partial x} = \log\left[\frac{2-x}{z_0} \frac{1-z_0}{x+\alpha-2}\right] + \frac{1}{x-2} + z_0'(x)\left[\frac{-x}{z_0(x)} + \frac{1-x}{1-z_0(x)}\right].
$$

Setting the above equal to zero and using

$$
z_0'(x)\left[\frac{1-x}{1-z_0(x)} - \frac{x}{z_0(x)}\right] = \frac{1}{2-x}
$$

we find that F is maximal when

$$
\frac{2-x}{\alpha} = z_0(x). \qquad (6.25)
$$

Note that this is the same as (6.22). Equation (6.25) defines $x = x_*(\alpha)$ and we set $z_*(\alpha) = z_0(x_*(\alpha))$. By using $(2-x)/(1-x) = 1/[\Delta(z_0)(1-z_0)]$ we can eliminate x in (6.25) and obtain

$$
\frac{1}{\alpha} - 1 = \frac{(1-z_*)\log(1-z_*)}{z_*}.
$$

The standard Laplace estimate of the integral in (6.24) is

$$
G(x_*;\alpha)\sqrt{\frac{2\pi}{j|F_{xx}(x_*;\alpha)|}} e^{jF(x_*;\alpha)}.
$$

Since $x_* = 2 - \alpha z_*$ we obtain

$$F(x_*; \alpha) = (1-\alpha)\log\alpha + (1-\alpha)\log(1-z_*) - \log z_*$$

so that

$$\tilde{F}_n^j \sim \frac{(2e)^{n-j}}{2\pi j^2} n^{j-n} z_*^{-j} (1-z_*)^{j-n} R(\alpha), \qquad n = j\alpha \tag{6.26}$$

where

$$R(\alpha) = G(x_*; \alpha)|F_{xx}(x_*; \alpha)|^{-1/2}.$$

By differentiating the equation defining z_* we obtain

$$z_*' = \frac{1}{\alpha}\frac{1-z_*}{(\alpha+1)z_* - 2} \tag{6.27}$$

where $z_*' = z_0'(x_*(\alpha))$. We also have

$$F_{xx} = \frac{1}{x-2} - \frac{1}{x+\alpha-2} - \left[\frac{1}{1-z_0} + \frac{1}{z_0}\right] z_0'$$

so that

$$|F_{xx}(x_*; \alpha)| = \left(\frac{1}{\alpha} + z_*'\right)\frac{1}{z_*(1-z_*)}. \tag{6.28}$$

Using (6.27) in (6.28) and setting $x_* = 2 - \alpha z_*$ we find that

$$R(\alpha) = \frac{1}{\alpha^{3/2}}\frac{z_*}{\sqrt{1-z_*}\sqrt{\alpha z_* - 1}}$$

and thus we have established part (c) of Lemma 2.

The above analysis holds for $\alpha > 2$ (i.e., $n > 2j$). For $1 < \alpha < 2$ we must truncate the lower limit of the sum in (6.17) at $m = 2j - 2 - n$, and thus the lower limit on the integral in (6.24) must be replaced by $2 - \alpha$. However, for any fixed $\alpha > 1$, the point $x_*(\alpha)$ satisfies $x_*(\alpha) > 2 - \alpha$ (since $0 < z_* < 1$), so that the leading term for $\tilde{F}_n^j$ applies for all $\alpha > 1$. However, as $\alpha \to 1^-$ the interval of integration shrinks to zero and we must reconsider the discrete sum in (6.17).

Next we consider $n \to \infty$ with $k = n - j = O(1)$. Truncating the sum in (6.17), setting $j - m - 2 = \ell$ and using part (1) of Lemma 1, we are led to

$$\tilde{F}_n^j \sim 2^k \sum_{\ell=0}^{k} \frac{1}{(k-\ell)!}\frac{1}{\ell!}\frac{1}{n-k-\ell-2}(\log(n-k-\ell-2))^\ell \sim \frac{2^k}{k!}\frac{[\log n]^k}{n},$$

since the dominant term in the sum is that with $\ell = k$. This yields part (e) of Lemma 2.

Finally we consider the limit $k, n \to \infty$ with $k = O(\log n)$. Now we use part (2) of Lemma 1 to obtain

$$\tilde{F}_n^j \sim 2^k \sum_{\ell=0}^{k} \frac{1}{(k-\ell)!}\frac{1}{(\ell+1)!}\frac{[\log(n-k-\ell-2)]^{\ell+1}}{n-k-\ell-2}\frac{1}{\Gamma\left(\frac{\ell+2}{\log(n-k-\ell-2)}\right)}.\exp\left(-\frac{\ell+2}{\log(n-k-\ell-2)}\right)$$

The major contribution to the sum will come from the upper limit, but now an infinite number of terms contribute to the asymptotic development. We use $(\ell+1)! = (k+1+\ell-k)! \sim (k+1)!k^{\ell-k}$ and obtain

$$\begin{aligned} \tilde{F}_n^j &\sim \frac{2^k}{n}\frac{e^{-\nu}}{\Gamma(\nu)}\frac{1}{(k+1)!}\sum_{\ell=0}^{k}\left(\frac{\log n}{k}\right)^{\ell-k}\frac{(\log n)^{k+1}}{(k-\ell)!} \\ &\sim \frac{2^k}{n}\frac{e^{-\nu}}{\Gamma(\nu)}\frac{k^{-k}e^k}{\sqrt{2\pi}k^{3/2}}(\log n)^{k+1}\sum_{i=0}^{\infty}\frac{\nu^i}{i!}, \qquad \nu = \frac{k}{\log n}. \end{aligned}$$

We have thus established part(d) of Lemma 2. ■

We now have a rather complete description of the asymptotic behavior of $\tilde{F}_n^j$. We next discuss the range of validity of the approximation $1 - L_n^k = 1 - L_n^{n-j} = F_n^j \sim \tilde{F}_n^j$. As discussed at the beginning of this section, the terms proportional to c_ℓ for $\ell \geq 2$ in (6.6) are negligible on both the j and α scales. In view of the estimates in Appendix A, this is no longer true in the near right tail, where $k = \nu \log n$ and $A < \nu < \infty$. We thus analyze this case separately in Section 7. Upon examining the five expressions in Lemma 2, we see that in cases (a)–(c) and (e), $\tilde{F}_n^j$ is asymptotically small. However, when $k = \nu \log n$ and ν lies in the range $A' < \nu < A$, where A and A' are the two solutions to $z\log(2e/z) = 1$ (thus $0 < A' < 1 < A = 4.311\ldots$), then $\tilde{F}_n^j$ is asymptotically large in n. This would lead to $L_n^k < 0$ and thus this range is clearly not in the right tail. Now suppose we *assume* that either (c) or (d) are valid in the near right tail. We can easily show that (c) and (d) asymptotically match in an intermediate limit where $\alpha \downarrow 1$ and $\nu \to \infty$. By letting $\nu \to \infty$ in (d) we obtain the solution in the matching region as

$$\tilde{F}_n^j \sim \frac{(2e)^k}{2\pi n k}\sqrt{\log n}\left(\frac{\log n}{k}\right)^k \exp\left(-\frac{k}{\log n}\log\left(\frac{k}{\log n}\right) + \frac{k}{\log n}\right). \tag{6.29}$$

Note that this becomes $O(1)$ in n when

$$k = A\log n - \frac{1}{2}\frac{A}{A-1}\log\log n + O(1).$$

Thus the assumption of being able to "push" the relation $F_n^j \sim \tilde{F}_n^j$ into the near right tail (i.e., the ν-scale) leads to the value of B in (2.9), which was conjectured to be true in the past. However, the analysis of the ν and ζ scales in Section 7 and Section 8 will exclude this possibility.

7 Near Right Tail

We consider $k = \nu \log n$, $n \to \infty$ and $\nu \in (A, \infty)$. Since we are still in the right tail we consider the linear problem (5.15) and change variables from (k, n) to (ν, n) with

$$F_n^j = U(\nu; n) = U\left(\frac{k}{\log n}; n\right) = U\left(\frac{n-j}{\log n}; n\right) \tag{7.1}$$

where $k = n - j$. It follows that

$$F_{n+1}^j = U\left(\nu \frac{\log n}{\log(n+1)} + \frac{1}{\log(n+1)}; n+1\right) \tag{7.2}$$

and

$$F_{n-\ell}^{j-\ell} = U\left(\nu \frac{\log n}{\log(n-\ell)}; n-\ell\right). \tag{7.3}$$

Using (7.1)–(7.3) in (5.15) yields

$$U\left(\nu \frac{\log n}{\log(n+1)} + \frac{1}{\log(n+1)}; n+1\right) = \frac{2}{n+1} \sum_{\ell=0}^{n-\nu \log n - 1} U\left(\nu \frac{\log n}{\log(n-\ell)}; \quad n-\ell\right). \tag{7.4}$$

We analyze (7.4) by a (linear) WKB expansion. That is, we assume an asymptotic expansion of the form

$$U(\nu; n) = e^{-a(\nu)\log n} (\log n)^\delta \left[b(\nu) + \frac{b^{(1)}(\nu)}{\log n} + \frac{b^{(2)}(\nu)}{\log^2 n} + O(\log^{-3} n)\right]. \tag{7.5}$$

Here δ is a constant which must be determined along with $a(\nu)$ and $b(\nu)$.

We shall show that in order to uniquely determine U (or even $a(\nu)$), we need the behavior of U as $\nu \to \infty$. We require that as $\nu \to \infty$ U asymptotically matches to the expansion on the α-scale, as $\alpha \downarrow 1$. The basic idea of *matched asymptotic expansion* is that an approximate solution to a given problem is sought not as a single expansion in terms of a single scale but as two or more separate expansions in terms of two or more scales, each of which is valid in part of the domain. We chose the scales so that the overall expansion covers the whole domain and that the domains of validity of neighboring expansions *overlap*. Because the domains overlap, we can *match* or blend the neighboring expansions. If this is possible, the resulting solution is called the *matched asymptotic expansion* (cf. [21]).

In our case, by rewriting (6.29) in terms of ν, this implies that

$$U(\nu; n) \sim \frac{(2e)^{\nu \log n}}{2\pi n \nu} \frac{\nu^{-\nu \log n}}{\sqrt{\log n}} e^{-\nu \log \nu + \nu}, \qquad \nu \to \infty. \tag{7.6}$$

We thus have

$$a(\nu) \sim \nu \log \nu - \nu \log(2e) + 1, \qquad \nu \to \infty \tag{7.7}$$

$$b(\nu) \sim \frac{1}{2\pi\nu} e^{\nu - \nu \log \nu}, \qquad \nu \to \infty \tag{7.8}$$

and $\delta = -1/2$. We also give an independent derivation of the value of δ below.

Using (7.5) and the expansions

$$\nu \frac{\log n}{\log(n+1)} + \frac{1}{\log(n+1)} = \nu + \frac{1}{\log n} + O(n^{-1}),$$

$$\frac{\log n}{\log(n-\ell)} = 1 - \frac{1}{\log n} \log\left(1 - \frac{\ell}{n}\right) + \frac{1}{\log^2 n} \log^2\left(1 - \frac{\ell}{n}\right) + O(\log^{-3} n)$$

we find that the left side of (7.4) becomes

$$(\log n)^\delta e^{-a(\nu)\log n} e^{-a'(\nu)} \left[1 - \frac{a''(\nu)}{2\log n}\right] \left[b(\nu) + \frac{b^{(1)}(\nu) + b'(\nu)}{\log n} + O(\log^{-2} n)\right]. \tag{7.9}$$

The right side of (7.4) is asymptotically equal to

$$\frac{2}{n}\sum_{\ell=0}^{n} (\log n)^\delta \left[1 + \delta \frac{\log(1-\ell/n)}{\log n} + O(\log^{-2} n)\right] \left[b(\nu) + \frac{1}{\log n}\left(b^{(1)}(\nu) - \nu b'(\nu) \log\left(1 - \frac{\ell}{n}\right)\right)\right]$$
$$\times e^{-a(\nu)\log n} e^{-\log(1-\ell/n)(a(\nu) - \nu a'(\nu))} \left[1 - \frac{\nu^2 a''(\nu)}{2\log n} \log^2\left(1 - \frac{\ell}{n}\right) + O(\log^{-2} n)\right].$$

Using the Euler-MacLaurin formula we find that the above is asymptotic to

$$\begin{aligned}(\log n)^\delta e^{-a(\nu)\log n} \quad &\times \quad \left(2b(\nu)\int_0^1 (1-x)^{\nu a'(\nu) - a(\nu)} dx\right. \\ &+ \quad \frac{1}{\log n}\left[2\int_0^1 (1-x)^{\nu a'(\nu)-a(\nu)} (b^{(1)}(\nu) - \nu b'(\nu)\log(1-x)\right. \\ &+ \quad \left.\left.\delta b(\nu)\log(1-x) - \frac{1}{2}\nu^2 a''(\nu) b(\nu)\log^2(1-x))dx\right] + O(\log^{-2} n)\right).\end{aligned} \tag{7.10}$$

We cancel the common factor $(\log n)^\delta e^{-a(\nu)\log n}$ in (7.9) and (7.10), compare terms of orders $O(1)$ and $O(\log^{-1} n)$, and explicitly evaluate the x-integrals in (7.10). This yields

$$\nu a'(\nu) - a(\nu) + 1 = 2e^{a'(\nu)} \tag{7.11}$$

and

$$\left[b'(\nu) - \frac{1}{2}b(\nu)a''(\nu)\right] e^{-a'(\nu)} = -\frac{2[\delta b(\nu) - \nu b'(\nu)]}{[\nu a'(\nu) - a(\nu) + 1]^2} - \frac{2\nu^2 a''(\nu) b(\nu)}{[\nu a'(\nu) - a(\nu) + 1]^3}. \tag{7.12}$$

The non-linear ODE (7.11) is the Clairaut equation (cf. [12], Chap. 2.45)). To solve it we differentiate (7.11) with respect to ν to find that

$$\nu a''(\nu) = 2e^{a'(\nu)} a''(\nu)$$

so that either $a''(\nu) = 0$ or $a'(\nu) = \log(\nu/2)$. The former leads to $a(\nu) = k\nu + k'$ and then (7.11) is satisfied if $k' = 1 - 2e^k$. Thus $a(\nu) = k\nu - 2e^k + 1$ is a one-parameter family of linear solutions. The solution with $a'(\nu) = \log(\nu/2)$ is the geometric envelope of this family and it is explicitly given by

$$a(\nu) = \nu \log \nu - \nu + 1 - \nu \log 2. \tag{7.13}$$

In view of (7.7) the linear solutions must be rejected. Note that

$$a(A) = 0 \quad \text{and} \quad a'(A) = \log\left(\frac{A}{2}\right) = 1 - \frac{1}{A} > 0. \tag{7.14}$$

Thus $a(\nu) > 0$ for $\nu > A$, and then (7.13) shows that (7.7) holds for all $\nu > A$, not just $\nu \to \infty$.

Using (7.13) we find that $\nu a'(\nu) - a(\nu) + 1 = \nu$ with which (7.12) becomes

$$\left[b'(\nu) - \frac{1}{2}b(\nu)\frac{1}{\nu}\right]\frac{2}{\nu} = -\frac{2}{\nu^2}[\delta b(\nu) - \nu b'(\nu)] - \frac{2}{\nu^2}b(\nu)$$

so that $b(\nu) = (2\delta + 2)b(\nu)$ and hence $\delta = -1/2$. This confirms the result we obtained via asymptotic matching. However, setting $\delta = -1/2$ in (7.12) and using (7.13) yields "0 = 0", and thus $b(\nu)$ cannot be determined! In Section 10 we briefly discuss the numerical computation of $b(\nu)$, but this too is problematic. In view of (7.8) we have the behavior as $\nu \to \infty$ and in Section 8 we discuss the behavior of $b(\nu)$ as $\nu \to A$.

Thus we have been able to determine δ and $a(\nu)$ in (7.5), but not the $O(1)$ factor $b(\nu)$. We also note that on the ν-scale $1 - L_n^k = O(n^{-a(\nu)})$, which is *algebraically* small in n. In this section we neglected the non-linear terms in (5.2). However, these are (roughly) of order $O(n^{-2a(\nu)})$ and since $a(\nu) > 0$ for $\nu > A$, they do not contribute to the asymptotic series in (7.5), whose terms are $O(n^{-a(\nu)}(\log n)^{-p})$ for $p = \frac{1}{2}, \frac{3}{2}, \dots$.

8 Central Region

We consider the important central region, where L_n^k undergoes the transition from $L_n^k \approx 0$ to $L_n^k \approx 1$, and where most of the mass is concentrated. Now we must analyze the full non-linear problem (2.2).

We define ζ by

$$\zeta = k - \bar{A}\log n - \bar{B}\log\log n \tag{8.1}$$

and for now we treat $\bar{A}$ and $\bar{B}$ as arbitrary parameters. They will ultimately be determined by asymptotically matching the central and near right tail regions. We set

$$L_n^k = f(\zeta; n) = f(k - \bar{A}\log n - \bar{B}\log\log n; n)$$

and note that

$$\begin{aligned} L_{n+1}^{k+1} &= f\left(\zeta + 1 - \bar{A}\log\left(1+\frac{1}{n}\right) - \bar{B}\log\left(\frac{\log(n+1)}{\log n}\right); n+1\right), \\ L_{\ell}^{k} &= f\left(\zeta - \bar{A}\log\left(\frac{\ell}{n}\right) - \bar{B}\log\left(\frac{\log \ell}{\log n}\right); \ell\right). \end{aligned}$$

Hence, in terms of ζ and n, (2.2) becomes

$$\begin{aligned} & f\left(\zeta + 1 - \bar{A}\log\left(1+\frac{1}{n}\right) - \bar{B}\log\left(\frac{\log(n+1)}{\log n}\right); n+1\right) \qquad (8.2) \\ = & \frac{1}{n+1}\sum_{\ell=0}^{n} f\left(\zeta - \bar{A}\log\left(\frac{\ell}{n}\right) - \bar{B}\log\left(\frac{\log \ell}{\log n}\right); \ell\right) f\left(\zeta - \bar{A}\log\left(1-\frac{\ell}{n}\right) - \bar{B}\log\left(\frac{\log(n-\ell)}{\log n}\right); n-\ell\right). \end{aligned}$$

We assume (as $n \to \infty$ with ζ fixed) an expansion of the form

$$f(\zeta; n) = f(\zeta) + \frac{1}{\log n} f^{(1)}(\zeta) + \frac{1}{\log^2 n} f^{(2)}(\zeta) + \cdots . \qquad (8.3)$$

Using (8.3) in (8.2), setting $x = \ell/n$ and approximating the sum by an integral leads to

$$f(\zeta+1) = \int_0^1 f(\zeta - \bar{A}\log x) f(\zeta - \bar{A}\log(1-x))dx \qquad (8.4)$$

and

$$\begin{aligned} & f^{(1)}(\zeta+1) - 2\int_0^1 f^{(1)}(\zeta - \bar{A}\log x) f(\zeta - \bar{A}\log(1-x))dx \qquad (8.5) \\ = & -2\bar{B}\int_0^1 (\log x) f'(\zeta - \bar{A}\log x) f(\zeta - \bar{A}\log(1-x))dx. \end{aligned}$$

Note that to obtain the equation (8.4) for the leading term, we can weaken (8.3) to the assumption that $f(\zeta; n) \to f(\zeta)$ as $n \to \infty$.

As $\zeta \to \infty$ we are entering the right tail so we expect that $f(\zeta) \to 1$ as $\zeta \to \infty$. Setting $f(\zeta) = 1 - g(\zeta)$ we obtain from (8.4)

$$g(\zeta+1) - 2\int_0^1 g(\zeta - \bar{A}\log x)dx = -\int_0^1 g(\zeta - \bar{A}\log x) g(\zeta - \bar{A}\log(1-x))dx. \qquad (8.6)$$

Furthermore we denote the solution to the linearized version of (8.6) as $g_L(\zeta)$, which satisfies

$$\begin{aligned} g_L(\zeta+1) &= 2\int_0^1 g_L(\zeta - \bar{A}\log x)dx \qquad (8.7) \\ &= 2\int_0^\infty g_L(\zeta + \bar{A}t)e^{-t}dt. \end{aligned}$$

We note that in (8.6) we have decomposed the non-linear integral operator into linear and non-linear parts, in a manner similar to (5.2).

We can easily solve (8.7) exactly. Indeed the equation admits exponential solutions of the form $e^{-a_0\zeta}$ provided that a_0 satisfies

$$e^{-a_0} = \frac{2}{1+a_0\bar{A}} > 0. \tag{8.8}$$

We thus consider $a_0 = a_0(\bar{A})$. By plotting the function $F(a_0) = (2e^{a_0} - 1)/a_0 = \bar{A}$ versus a_0, we see that if $\bar{A} > A = 4.311\ldots$, then (8.8) has two solutions, which we denote by a_+ and a_- with $a_+(\bar{A}) > a_-(\bar{A})$. When $\bar{A} < A$, (8.8) has no solution and when $\bar{A} = A$ then (8.8) has the unique solution $a_0 = 1 - 1/A$, which is a double root of (8.8). We have thus obtained

$$g_L(\zeta) = \begin{cases} c_+e^{-a_+(\bar{A})\zeta} + c_-e^{-a_-(\bar{A})\zeta}, & \bar{A} > A \\ (c_1\zeta + \tilde{c})\exp\left[-\left(1-\frac{1}{A}\right)\zeta\right], & \bar{A} = A = 4.311\ldots \\ 0, \qquad \bar{A} < A \end{cases} . \tag{8.9}$$

Here $c_1, \tilde{c}, c_\pm$ are arbitrary constants.

The linearized problem (8.7) also admits complex exponential solutions. However, in Appendix C we show, by a combination of analytic and numerical studies, that such solutions are not relevant to the present problem. Thus we restrict ourselves to real roots of (8.8).

Next we determine $\bar{A}$ and $\bar{B}$ by asymptotic matching. On the ν-scale $1 - L_n^k$ is given by (7.5) with (7.13) and $\delta = -1/2$. Near $\nu = A$ we have

$$a(\nu) = a'(A)(\nu - A) + \frac{1}{2}a''(A)(\nu - A)^2 + \cdots = \left(1 - \frac{1}{A}\right)(\nu - A) + \frac{1}{2A}(\nu - A)^2 + \cdots .$$

But, $\nu - A = (k - A\log n)/\log n = (\zeta + (\bar{A} - A)\log n + \bar{B}\log\log n)/(\log n)$. Thus as $\nu \downarrow A$ the expansion on the ν-scale has the local behavior

$$b(A)(\log n)^{-1/2}\exp\left[-\left(1-\frac{1}{A}\right)\zeta - \left(1-\frac{1}{A}\right)(\bar{A}-A)\log n - \left(1-\frac{1}{A}\right)\bar{B}\log\log n\right] \tag{8.10}$$

if $b(A) > 0$, or

$$\begin{aligned} b'(A)[\zeta \;\; &+ \;\; (\bar{A}-A)\log n + \bar{B}\log\log n](\log n)^{-3/2} \\ &\times \;\; \exp\left[-\left(1-\frac{1}{A}\right)(\zeta + (\bar{A}-A)\log n + \bar{B}\log\log n)\right] \end{aligned} \tag{8.11}$$

if $b(A) = 0$ and $b'(A) \neq 0$. Since clearly $g(\zeta) \sim g_L(\zeta)$ as $\zeta \to \infty$, the asymptotic matching of the ζ and ν-scales implies that $\bar{A} = A = 4.311\ldots$. Thus the leading term for $\mathbf{E}[\mathcal{H}_n]$ is $A\log n$, as is well-known. The conclusion that $\bar{A} = A$ is based only on comparing the exponential growth rates in (8.9) and (8.10), (8.11). It follows that $g(\zeta) \sim (c_1\zeta +$

$\tilde{c})\exp\left[-\left(1-\frac{1}{A}\right)\zeta\right]$ as $\zeta\to\infty$. By comparing the algebraic factors we see that there are two possibilities:

$$\begin{array}{llll} (1) & c_1=0 \;\Rightarrow\; \tilde{c}=b(A) & \text{and} & \bar{B}=-\frac{1}{2}\frac{A}{A-1} \\ (2) & c_1\neq 0 \;\Rightarrow\; c_1=b'(A) & \text{and} & \bar{B}=-\frac{3}{2}\frac{A}{A-1}\equiv B. \end{array} \tag{8.12}$$

Note that in the former case $(\log n)^{-1/2}\exp\left[-\left(1-\frac{1}{A}\right)\bar{B}\log\log n\right]=1$ and in the latter case $(\log n)^{-3/2}\exp\left[-\left(1-\frac{1}{A}\right)\bar{B}\log\log n\right]=1$.

We now show that case (1) in (8.12) leads to a contradiction. If $g(\zeta)$ behaved purely as an exponential as $\zeta\to\infty$, then we can construct the exact solution to the full non-linear problem in the form

$$g(\zeta)=\sum_{m=1}^{\infty}e(m)e^{-ma_0\zeta},\qquad a_0=1-\frac{1}{A}. \tag{8.13}$$

Note that this corresponds to solving (8.6) by the method of successive approximations, with the initial guess being the exponential $e^{-a_0\zeta}$. Once we conclude that $g(\zeta)\sim e(1)e^{-a_0\zeta}$ as $\zeta\to\infty$, we can set $g(\zeta)-e(1)e^{-a_0\zeta}=g_*(\zeta)$ in (8.6) to conclude that $g_*(\zeta)=O(e^{-2a_0\zeta})$ as $\zeta\to\infty$, and so on. Using (8.13) in (8.6) leads to

$$\begin{aligned} \sum_{m=1}^{\infty}e(m)e^{-ma_0}e^{-ma_0\zeta} \;&-\; 2\sum_{m=1}^{\infty}e(m)\int_0^1 e^{-ma_0\zeta}x^{ma_0A}dx \\ &= -\sum_{m,\ell\geq 1}e(m)e(\ell)\int_0^1 e^{-(m+\ell)a_0\zeta}x^{ma_0A}(1-x)^{\ell a_0A}dx. \end{aligned} \tag{8.14}$$

By comparing coefficients of $e^{-ma_0\zeta}$, using

$$\int_0^1 x^{ma_0A}(1-x)^{\ell a_0A}dx=B(1+ma_0A,1+\ell a_0A)=\frac{\Gamma(m(A-1)+1)\Gamma(\ell(A-1)+1)}{\Gamma((m+\ell)(A-1)+2)},$$

setting

$$e(N)=\frac{\bar{e}(N)}{\Gamma(N(A-1)+1)},$$

and noting that $e^{-a_0}=2/A$, we are led to the recurrence

$$\bar{e}(N)\left[2-(N(A-1)+1)\left(\frac{2}{A}\right)^N\right]=\sum_{\ell=1}^{N-1}\bar{e}(\ell)\bar{e}(N-\ell) \tag{8.15}$$

for $N\geq 2$. By the matching condition we have $e(1)=\tilde{c}=b(A)$ so we set

$$\bar{e}(N)=[b(A)]^N[\Gamma(A)]^N d(N)$$

to find that $d(N)$ satisfies, for $N\geq 2$

$$d(N)\left[2-(N(A-1)+1)\left(\frac{2}{A}\right)^N\right]=\sum_{\ell=1}^{N-1}d(\ell)d(N-\ell),\quad d(1)=1. \tag{8.16}$$

In terms of $d(N)$ we have

$$g(\zeta) = \sum_{m=1}^{\infty} \frac{(b(A)\Gamma(A))^m}{\Gamma(m(A-1)+1)} d(m) \exp\left[-m\left(1-\frac{1}{A}\right)\zeta\right]. \tag{8.17}$$

From (8.16) it follows that $d(N) > 0$ for all $N \geq 1$.

In Appendix B we show by a combination of analytical and numerical methods that

$$d(N) \sim L_2 N^{-3/2} (L_1)^N, \qquad N \to \infty \tag{8.18}$$

where L_1 and L_2 are constants, whose approximate numerical values are

$$L_1 \approx 3.89, \qquad L_2 \approx .377. \tag{8.19}$$

But, in view of (8.18) we see that the coefficients in the series (8.17) decay faster than geometrically, roughly as $\exp(-m(A-1)\log m)$. It follows that (8.17) defines an *entire* function of ζ. But then (since $d(m) > 0$ for all m) $g(\zeta) \to \infty$ as $\zeta \to -\infty$ and thus $f(\zeta) \to -\infty$ as $\zeta \to -\infty$. Since this is unacceptable for a probability distribution, we conclude that we cannot have $c_1 = 0$ in (8.12) and thus case (2) must hold.

We observe that (8.4) is independent of $\bar{B}$. The solution to this equation is not unique, for if $f_0(\zeta)$ is a solution, so is $f_0(\zeta + c)$ for any constant c. The asymptotic matching condition $1 - f(\zeta) \sim c_1 \zeta \exp\left[-\left(1-\frac{1}{A}\right)\zeta\right]$, $\zeta \to \infty$ is needed to uniquely specify $f(\zeta)$. Setting $\bar{A} = A$ we can recast this equation as follows. We let

$$\Delta = \frac{1}{A} e^{-\zeta/A}, \quad \zeta = -A\log(A\Delta), \quad f(\zeta) = \bar{f}(\Delta).$$

Then (8.4) becomes

$$\begin{aligned} \bar{f}\left(\frac{A}{2e}\Delta\right) &= \int_0^1 \bar{f}(x\Delta)\bar{f}((1-x)\Delta)dx \\ &= \frac{1}{\Delta}\int_0^{\Delta} \bar{f}(y)\bar{f}(\Delta - y)dy. \end{aligned} \tag{8.20}$$

Introducing the Laplace transform

$$F(\eta) = \int_0^{\infty} e^{-\Delta\eta} \bar{f}(\Delta) d\Delta$$

we find from (8.20) that

$$-e^{1/A} \frac{d}{d\eta}[F(e^{1/A}\eta)] = F^2(\eta)$$

or

$$-F'(\eta) = e^{-2/A} F^2(\eta e^{-1/A}). \tag{8.21}$$

This is precisely the equation analyzed by Drmota [8] who proved that it has a proper solution. In terms of $F(\eta)$ we thus have

$$f(\zeta) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \exp\left(\frac{\eta}{A}e^{-\zeta/A}\right) F(\eta)d\eta \tag{8.22}$$

for $a > 0$ in the complex η-plane. Note that the solution to (8.21) is also not unique: if $F_0(\eta)$ is a solution, so is $\lambda F_0(\lambda\eta)$ for any λ. We also note that $\bar{f}(0) = f(\infty) = 1$ implies

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} F(\eta)d\eta = 1.$$

The contractive nature of (8.21) implies that $F(\eta)$ is an entire function, and this was rigorously shown by Drmota in [8]. The solution to (8.21) can be made unique by using a normalization such as $F(0) = 1$. However, such an arbitrary choice would not give the proper normalization (or shift) for $f(\zeta)$; for this we needed the asymptotic matching condition.

We can get some information about the behavior of $F(\eta)$ as $\eta \to \pm\infty$, and that of $f(\zeta)$ as $\zeta \to -\infty$. For $\eta \to +\infty$, it is shown in [8] that any (non-zero) solution to (8.21) satisfies $F(\eta) \sim 1/\eta$ as $\eta \to \infty$. This can be extended to a full asymptotic series of the form

$$F(\eta) \sim \frac{1}{\eta} + \sum_{m=1}^{\infty} \frac{1}{\eta^{1+m(A-1)}}\mathcal{P}_m(\log\eta) \tag{8.23}$$

where $\mathcal{P}_m$ is a polynomial of degree m. Using (8.23) in (8.21) and also using $\exp\left(1-\frac{1}{A}\right) = \frac{A}{2}$ we are led to the recurrence

$$\begin{aligned}[1+m(A-1)]\mathcal{P}_m(\log\eta) - \mathcal{P}_m'(\log\eta) - 2\left(\frac{A}{2}\right)^m \mathcal{P}_m\left(\log\eta - \frac{1}{A}\right) \quad (8.24)\\ = \left(\frac{A}{2}\right)^m \sum_{\ell=1}^{m-1} \mathcal{P}_\ell\left(\log\eta - \frac{1}{A}\right)\mathcal{P}_{m-\ell}\left(\log\eta - \frac{1}{A}\right), \quad m \geq 1.\end{aligned}$$

Upon setting $m = 1$ in (8.24) we obtain

$$A\mathcal{P}_1(X) - \mathcal{P}_1'(X) - A\mathcal{P}_1\left(X - \frac{1}{A}\right) = 0.$$

This is satisfied by an arbitrary linear polynomial, so we write

$$\mathcal{P}_1(X) = DX + \tilde{D}. \tag{8.25}$$

Then with $m = 2$, (8.24) yields

$$(2A-1)\mathcal{P}_2(X) - \mathcal{P}_2'(X) - \frac{A^2}{2}\mathcal{P}_2\left(X - \frac{1}{A}\right) = \left(\frac{A}{2}\right)^2 (DX + \tilde{D})^2. \tag{8.26}$$

We thus obtain

$$\mathcal{P}_2(X) = \ell_2 X^2 + \ell_1 X + \ell_0$$

where

$$\begin{aligned}
\ell_2 &= \frac{1}{2}\frac{A^2}{4A-2-A^2}D^2,\\
\ell_1 &= \frac{A^2}{4A-2-A^2}D\left[\frac{2-A}{4A-2-A^2}D+\tilde{D}\right],\\
\ell_0 &= \frac{A^2}{4A-2-A^2}\left[\frac{\tilde{D}^2}{2}+\frac{(2-A)D\tilde{D}}{4A-2-A^2}+\frac{D^2}{2(4A-2-A^2)}+\frac{(2-A)^2D^2}{(4A-2-A^2)^2}\right].
\end{aligned}$$

We use (8.23) in (8.22). Since $F(\eta)$ is entire we can shift the contour far to the right, where $\eta \to +\infty$. It follows that as $\zeta \to \infty$

$$f(\zeta) \sim 1 + \sum_{m=1}^{\infty} \frac{1}{2\pi i}\int_{M-i\infty}^{M+i\infty} \frac{\mathcal{P}_m(\log\eta)}{\eta^{1+m(A-1)}} \exp\left[\eta\frac{e^{-\zeta/A}}{A}\right] d\eta \tag{8.27}$$

for arbitrary large $M > 0$, and thus

$$f(\zeta) \sim 1 - \sum_{m=1}^{\infty} \bar{\mathcal{P}}_m(\zeta)\exp\left[-m\zeta\left(1-\frac{1}{A}\right)\right]$$

where $\bar{\mathcal{P}}_m(\cdot)$ are polynomials of degree m, which can be calculated from $\mathcal{P}_m(\cdot)$ as follows:

$$\bar{\mathcal{P}}_m(\zeta) = -\frac{A^{-m(A-1)}}{2\pi i}\int_{M-i\infty}^{M+i\infty}\frac{e^w}{w^{1+m(A-1)}}\mathcal{P}_m\left(\frac{\zeta}{A}+\log A+\log w\right)dw.$$

By contour integration we have

$$\begin{aligned}
\frac{1}{2\pi i}\int_{M-i\infty}^{M+i\infty}\frac{e^w}{w^A}dw &= \frac{1}{\Gamma(A)},\\
\frac{1}{2\pi i}\int_{M-i\infty}^{M+i\infty}\frac{e^w\log w}{w^A}dw &= \frac{\Gamma'(A)}{\Gamma^2(A)}
\end{aligned}$$

so that, in view of (8.25), we obtain

$$\bar{\mathcal{P}}_1(\zeta) = -\frac{A^{1-A}}{\Gamma(A)}\left[D\left(\frac{\zeta}{A}+\log A+\frac{\Gamma'(A)}{\Gamma(A)}\right)+\tilde{D}\right].$$

In view of (8.9), the constants $(c_1, \tilde{c})$ are related to $(D, \tilde{D})$ via

$$c_1 = -\frac{A^{-A}}{\Gamma(A)}D, \quad \tilde{c} = -\frac{A^{1-A}}{\Gamma(A)}\left[D\left(\log A+\frac{\Gamma'(A)}{\Gamma(A)}\right)+\tilde{D}\right]$$

so we clearly have $D < 0$.

The numerical studies in Section 10 show that $F(\eta)$ grows rapidly as $\eta \to -\infty$. We can estimate this growth by seeking an asymptotic solution to (8.21) in the form

$$F(\eta) \sim \nu_4 |\eta|^{\nu_3} \exp(\nu_2 |\eta|^{\nu_1}), \qquad \eta \to -\infty.$$

Using the above in (8.21) and retaining only the leading term for $F'(\eta)$ leads to the balance

$$\nu_1 \nu_2 \nu_4 |\eta|^{\nu_1+\nu_3-1} \exp(\nu_2|\eta|^{\nu_1}) \sim e^{-2/A} \nu_4^2 |\eta|^{2\nu_3} \exp\left(-\frac{2}{A}\nu_3\right) \exp(2\nu_2 |\eta|^{\nu_1} e^{-\nu_1/A})$$

so we must have

$$\nu_1 = A \log 2, \quad \nu_3 = A \log 2 - 1, \quad \nu_4 = (4A \log 2)\nu_2$$

and ν_2 is arbitrary. We have thus formally obtained

$$F(\eta) \sim 4A(\log 2)\nu_2(-\eta)^{A \log 2 - 1} \exp[\nu_2(-\eta)^{A \log 2}], \quad \eta \to \infty. \tag{8.28}$$

The numerical calculation of ν_2 is discussed in Section 10.

We next analyze (8.4) (with $\bar{A} = A$) in the limit $\zeta \to -\infty$. Now we are getting into the left tail so we have $f(\zeta) \to 0^+$ as $\zeta \to -\infty$. We set $\xi = e^{-\zeta}$ with $f(\zeta) = h(e^{-\zeta}) = h(\xi) = f(-\log \xi)$ and obtain from (8.4)

$$h\left(\frac{\xi}{e}\right) = \int_0^1 h(\xi x^A) h(\xi(1-x)^A) dx. \tag{8.29}$$

As $\zeta \to -\infty$ we have $\xi \to +\infty$. We seek an asymptotic solution to (8.29) in the form

$$h(\xi) \sim \beta_2 \xi^{\beta_1} \exp[-c\xi^\beta].$$

We thus obtain

$$\begin{aligned} \beta_2 \left(\frac{\xi}{e}\right)^{\beta_1} \exp(-ce^{-\beta}\xi^\beta) \quad &\sim \quad \int_0^1 \beta_2^2 \xi^{2\beta_1} x^{A\beta_1} (1-x)^{A\beta_1} \exp[-c\xi^\beta (x^{A\beta} + (1-x)^{A\beta})] dx \\ &\sim \quad \beta_2^2 \xi^{2\beta_1} \left(\frac{1}{4}\right)^{A\beta_1} \sqrt{\frac{2\pi}{c\xi^\beta}} \frac{2^{A\beta/2} 2^{-3/2}}{\sqrt{A\beta(A\beta-1)}} \exp[-c\xi^\beta 2^{1-\beta A}] \end{aligned}$$

where the last step involved Laplace's method to evaluate the integral as $\xi \to \infty$ (the major contribution comes from $x \approx \frac{1}{2}$). It follows that

$$e^{-\beta} = 2^{1-\beta A} \quad \text{or} \quad \beta = \frac{\log 2}{A \log 2 - 1} = .3486\ldots\ .$$

Then we get $\beta_1 = \beta/2$ and

$$\beta_2 = 2\sqrt{\frac{2c}{\pi}} \sqrt{A\beta(A\beta-1)} = 2\sqrt{\frac{2c}{\pi}} \frac{\sqrt{A \log 2}}{A \log 2 - 1}.$$

In terms of ζ we have

$$f(\zeta) \sim 2\sqrt{\frac{2c}{\pi}}\frac{\sqrt{A\log 2}}{A\log 2 - 1}e^{-\beta\zeta/2}\exp(-ce^{-\beta\zeta}), \quad \zeta \to -\infty. \tag{8.30}$$

for some constant c. The value of c cannot be determined solely from (8.4), in view of the translation invariance of this equation.

An alternate argument for (8.30) is as follows. Since $F(\eta)$ is an entire function we shift the contour in (8.22) toward the left where we may use the expansion (8.28), that applies for $\eta \to -\infty$. This yields

$$f(\zeta) \sim \frac{1}{2\pi i}(4A\nu_2\log 2)\int_{-M-i\infty}^{-M+i\infty}(-\eta)^{A\log 2-1}\exp\left[\frac{\eta}{A}e^{|\zeta|/A} + \nu_2(-\eta)^{A\log 2}\right]d\eta$$

where $M > 0$ is a large positive number. For $\zeta \to -\infty$ the above has a saddle point where

$$-\eta = \eta_* \equiv \left[\frac{e^{-\zeta/A}}{\nu_2 A^2\log 2}\right]^{\frac{1}{A\log 2-1}}.$$

Then the standard saddle point approximation leads to

$$\begin{aligned} f(\zeta) \;&\sim\; \frac{4A\nu_2\log 2}{\sqrt{2\pi}}\frac{e^{-\zeta/A}}{\nu_2 A^2\log 2}\frac{1}{\sqrt{\nu_2 A\log 2(A\log 2-1)}} \\ &\times\; \left[\frac{e^{-\zeta/A}}{\nu_2 A^2\log 2}\right]^{-\frac{A\log 2-2}{2(A\log 2-1)}}\exp\left[\frac{1-A\log 2}{A^2\log 2}\left(\frac{1}{\nu_2 A^2\log 2}\right)^{\frac{1}{A\log 2-1}}\exp\left(-\frac{\zeta\log 2}{A\log 2-1}\right)\right]. \end{aligned}$$

The above agrees precisely with (8.30) if ν_2 and c are related by

$$c = (A\log 2-1)(A^2\log 2)^{-\frac{A\log 2}{A\log 2-1}}\nu_2^{-\frac{1}{A\log 2-1}}.$$

Having derived the behavior of $f(\zeta)$ as $\zeta \to -\infty$, we investigate the asymptotic matching between the ω and ζ scales. We recall that $\log_2\omega = \log_2 n - k$ and thus

$$\begin{aligned} \zeta &= \left(\frac{1}{\log 2} - A\right)\log n - B\log\log n - \log_2\omega, \\ \omega &= n^{1-A\log 2}e^{-\zeta\log 2}(\log n)^{-B\log 2}. \end{aligned}$$

By asymptotic matching, (8.30) should agree with the behavior of $\sqrt{n}A(\omega)e^{-n\phi(\omega)}$ as $\omega \to 0^+$. This is true provided that as $\omega \to 0$

$$\phi(\omega) \sim \omega^{\frac{1}{A\log 2-1}}(-\log\omega)^{\beta B}\hat{c}, \quad \hat{c} = c(A\log 2-1)^{-\beta B}.$$

and

$$A(\omega) \sim \sqrt{\frac{8}{\pi}}\frac{\sqrt{A\log 2}}{A\log 2-1}\sqrt{\hat{c}}\,\omega^{\frac{1}{2}\frac{1}{A\log 2-1}}(-\log\omega)^{\beta B}.$$

Note that the above is consistent with the relationship between $\phi(\omega)$ and $A(\omega)$ in (4.8).

To summarize, we have shown that analysis of the central region involves solving the non-linear integral equation in (8.4). Our formal asymptotic results show that the value of B cannot be that in (2.9), and lead to the new conjecture in (2.7). Some numerical results to support this are presented in Section 10. We have also obtained the tails of $f(\zeta)$, as $\zeta \to \pm\infty$. Note that an extreme value (or double exponential) distribution would correspond to a solution to (8.4) in the form $\exp(-K_2 e^{-K_1\zeta})$. While this does not satisfy (8.4), our analysis shows that the distribution behaves (roughly) as a double exponential at both $\pm\infty$. However, the values of K_1 are different in these two limits. The analysis as $\zeta \to \infty$ would be consistent with $K_1 = 1 - 1/A = .768\ldots$, while the $\zeta \to -\infty$ result would lead to $K_1 = \beta = .348\ldots$

9 Numerical Studies of the Tails

We assess the numerical accuracy of our results for the M, j and α scales, and also determine the hitherto unknown function $\phi(\omega)$, which arose in the analysis of the ω scale.

We first consider $M = O(1)$, where we have shown in Section 3 that

$$L^k_{2^k-M} \sim \frac{2^{kM}}{(M-1)!} 4 \left(\frac{1}{2\sqrt{2}K_0} \right)^{2^k} \equiv L^k_n(\text{asy, } M\text{-scale}). \tag{9.1}$$

In Table 1 we compare (9.1) to the exact L^k_n for $1 \le M \le 5$, and $k = 4, 6, 8$. We see that the agreement is overall quite good. For a fixed M the result improves rapidly with k, and larger values of k allow for larger values of M to be used. The data in Table 1 are consistent with an error term in (9.1) of the form $1 + O(2^{-k})$, with the coefficient of the $O(2^{-k})$ correction term growing with M. Note also that if $k = 8$ and $M = O(1)$, the probability distribution is very small, of the order of 10^{-100}.

In Table 2 we consider the far right tail, where $j = n - k = O(1)$ and we have shown that

$$1 - L^k_n \sim \frac{2^n}{n!} \frac{n^{2j-2}}{(j-1)!} 2^{1-2j} \equiv 1 - L^k_n(\text{asy, } j\text{-scale}). \tag{9.2}$$

Note that (9.2) is exact for $j = 1$ and we may approximate $n!$ by Stirling's formula. When $j = 4$ and $n = 20$ the asymptotic result overestimates the true value by about 50%, but this improves to an error of about 4% when $j = 4$ and $n = 100$.

In Table 3 we compare the exact values of $1 - L^k_n$ to the asymptotic result on the α-scale. This corresponds to case (v) in Section 2 and applies when k is a significant fraction of n (e.g., $k = n/2$). In Table 3 we also give the uniform right tail result, which is $1 - L^k_n \sim \tilde{F}^j_n$

Table 1: Far Left Tail Comparison

k	M	L_n^k (exact)	L_n^k (asy, M-scale)
4	1	$.1680(10^{-4})$	$.1716(10^{-4})$
	2	$.2520(10^{-3})$	$.2745(10^{-3})$
	3	$.1764(10^{-2})$	$.2196(10^{-2})$
	4	$.7937(10^{-2})$	$.1171(10^{-1})$
	5	$.2646(10^{-1})$	$.4685(10^{-1})$
6	1	$.1315(10^{-23})$	$.1322(10^{-23})$
	2	$.8283(10^{-22})$	$.8459(10^{-22})$
	3	$.2568(10^{-20})$	$.2707(10^{-20})$
	4	$.5232(10^{-19})$	$.5774(10^{-19})$
	5	$.7896(10^{-18})$	$.9239(10^{-18})$
8	1	$.7265(10^{-102})$	$.7275(10^{-102})$
	2	$.1853(10^{-99})$	$.1862(10^{-99})$
	3	$.2353(10^{-97})$	$.2384(10^{-97})$
	4	$.1984(10^{-95})$	$.2034(10^{-95})$
	5	$.1251(10^{-93})$	$.1302(10^{-93})$

Table 2: Far Right Tail Comparison

n	j	$1-L_n^k$ (exact)	$1-L_n^k$ (asy, j-scale)
20	1	$.2155(10^{-12})$	$.2155(10^{-12})$
	2	$.2047(10^{-10})$	$.2155(10^{-10})$
	3	$.9056(10^{-9})$	$.1077(10^{-8})$
	4	$.2477(10^{-7})$	$.3592(10^{-7})$
100	1	$.6792(10^{-128})$	$.6792(10^{-128})$
	2	$.1681(10^{-124})$	$.1698(10^{-124})$
	3	$.2052(10^{-121})$	$.2122(10^{-121})$
	4	$.1648(10^{-118})$	$.1769(10^{-118})$
200	1	$.1019(10^{-314})$	$.1019(10^{-314})$
	2	$.1014(10^{-310})$	$.1019(10^{-310})$
	3	$.5009(10^{-307})$	$.5094(10^{-307})$
	4	$.1639(10^{-303})$	$.1698(10^{-303})$

where $\tilde{F}_n^j$ is given by (6.8) $\left(\text{with } \bar{c}_1 = \frac{1}{2}\right)$. The analysis in Section 6 showed that this applies on both the j and α scales, though this expression is more complicated in form (and also more difficult to evaluate numerically) than the j and α scale results. We consider $n = 20$ and $n = 100$, for various $k = O(n)$. These results show that while the α-scale expansion is reasonably accurate, the URT result always improves upon it. In Table 4 and Table 5 we consider $n = 20$ and $n = 100$, respectively, and compare $1 - L_n^k$ to the URT approximation. We decrease k until (the asymptotic) $1 - L_n^k$ exceeds 1; by then we are clearly out of the right tail. We see that as long as $1 - L_n^k$ is small, URT is reasonably accurate. However, if we define the "numerical near right tail" as those values of k where $10^{-3} \le 1 - L_n^k \le 10^{-2}$, then we see that the results in Table 4 ($n = 20$) are somewhat more accurate than those in Table 5 ($n = 100$). This is also consistent with the asymptotic analysis, which predicts that URT ceases to be valid in the near right tail, where $n \to \infty$ with $k = \nu \log n$, $A < \nu < \infty$. Larger values of n allow for a clearer resolution of this asymptotic range.

We next consider the ω-scale, where we have $L_n^k \sim \sqrt{n} A(\omega) e^{-n\phi(\omega)}$, but $\phi(\omega)$ is known only asymptotically for $\omega \to 0$ and $\omega \to 1$. We define

$$\phi_{\text{num}}(\omega; k) = \frac{1}{2n} \log n - \frac{1}{n} \log(L_n^k), \qquad \omega = n2^{-k} \tag{9.3}$$

Table 3: Right Tail Comparison

n	k	$1 - L_n^k$ (exact)	$1 - L_n^k$ (asy, α-scale)	$1 - L_n^k$ (URT)
20	15	$.4691(10^{-6})$	$.4927(10^{-6})$	$.4703(10^{-6})$
	10	$.1863(10^{-1})$	$.2065(10^{-1})$	$.1948(10^{-1})$
100	90	$.3965(10^{-103})$	$.4018(10^{-103})$	$.3965(10^{-103})$
	80	$.1444(10^{-81})$	$.1459(10^{-81})$	$.1445(10^{-81})$
	70	$.4650(10^{-63})$	$.4697(10^{-63})$	$.4652(10^{-63})$
	60	$.7168(10^{-47})$	$.7251(10^{-47})$	$.7176(10^{-47})$
	50	$.9548(10^{-33})$	$.9694(10^{-33})$	$.9576(10^{-33})$
	40	$.1211(10^{-20})$	$.1239(10^{-20})$	$.1220(10^{-20})$
	30	$.1087(10^{-10})$	$.1136(10^{-10})$	$.1112(10^{-10})$
	25	$.1275(10^{-6})$	$.1368(10^{-6})$	$.1333(10^{-6})$
	20	$.2923(10^{-3})$	$.3340(10^{-3})$	$.3231(10^{-3})$
	15	$.8922(10^{-1})$	.1257	.1201

Table 4: URT Approximation with $n = 20$

k	$1 - L_n^k$ (exact)	$1 - L_n^k$ (URT)
18	$.2047(10^{-10})$	$.2047(10^{-10})$
17	$.9056(10^{-9})$	$.9059(10^{-9})$
16	$.2477(10^{-7})$	$.2480(10^{-7})$
15	$.4691(10^{-6})$	$.4703(10^{-6})$
14	$.6527(10^{-5})$	$.6559(10^{-5})$
13	$.6905(10^{-4})$	$.6968(10^{-4})$
12	$.5673(10^{-3})$	$.5762(10^{-3})$
11	$.3661(10^{-2})$	$.3759(10^{-2})$
10	$.1863(10^{-1})$	$.1948(10^{-1})$
9	$.7455(10^{-1})$	$.8037(10^{-1})$
8	.2311	.2635
7	.5368	.6815

Table 5: URT Approximation with $n = 100$

k	$1 - L_n^k$ (exact)	$1 - L_n^k$ (URT)
98	$.1681(10^{-124})$	$.1681(10^{-124})$
95	$.9787(10^{-116})$	$.9787(10^{-116})$
90	$.3965(10^{-103})$	$.396510(10^{-103})$
80	$.1444(10^{-81})$	$.1445(10^{-81})$
70	$.4650(10^{-63})$	$.4652(10^{-63})$
60	$.7168(10^{-47})$	$.7176(10^{-47})$
50	$.9548(10^{-33})$	$.9576(10^{-33})$
40	$.1211(10^{-20})$	$.1220(10^{-20})$
30	$.1087(10^{-10})$	$.1112(10^{-10})$
25	$.1275(10^{-6})$	$.1333(10^{-6})$
20	$.2923(10^{-3})$	$.3231(10^{-3})$
19	$.1099(10^{-2})$	$.1240(10^{-2})$
18	$.3802(10^{-2})$	$.4405(10^{-2})$
17	$.1203(10^{-1})$	$.1444(10^{-1})$
16	$.3457(10^{-1})$	$.4352(10^{-1})$
15	$.8922(10^{-1})$	.1201
14	.2036	.3022
13	.4016	.6897

Table 6: Numerical Evaluation of the ϕ function.

(a)	k	$\phi^{(1)}_{\text{num}}$	ϕ_{num}
	5	.1997	.2863
	6	.2291	.2832
	7	.2485	.2810
	8	.2607	.2797

(b)	k	$\phi^{(1)}_{\text{num}}$	ϕ_{num}
	4	.7330	.8232
	5	.8201	.8755
	6	.8728	.9057
	7	.9041	.9232
	8	.9223	.9332

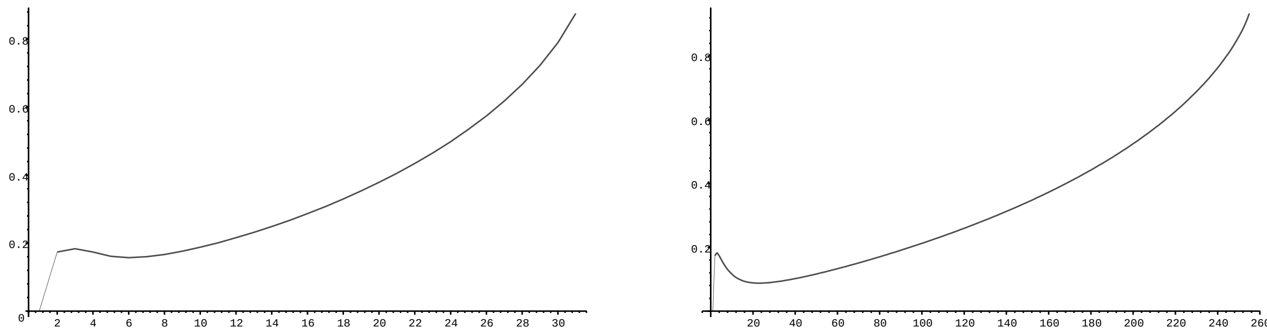
and

$$\phi^{(1)}_{\text{num}}(\omega;k) = -\frac{1}{n}\log(L_n^k) \tag{9.4}$$

where L_n^k is understood to be the exact (numerical) value. According to our WKB expansion we should have both ϕ_{num} and $\phi^{(1)}_{\text{num}} \to \phi(\omega)$ as k (or n) $\to \infty$ with ω held fixed. The convergence should be faster for ϕ_{num}, since (9.3) makes use of the algebraic factor $n^\gamma = \sqrt{n}$ in (4.4). In Table 6(a) we fix $\omega = 1/2$ and give ϕ_{num} and $\phi^{(1)}_{\text{num}}$ for $5 \le k \le 8$. Both sequences appear to converge to a value $\phi\left(\frac{1}{2}\right) \approx .28$, and the convergence is definitely faster for ϕ_{num}. In Table 6(b) we set $n = 2^k - 1$ so that $M = 1$ and $\omega = 1 - 2^{-k}$. Note that now we have the exact theoretical value $\phi(1) = c_* = \log(2\sqrt{2}K_0) = .945755\ldots$. We again give ϕ_{num} and $\phi^{(1)}_{\text{num}}$, for $4 \le k \le 8$. Both sequences again appear to converge with ϕ_{num} being the more rapidly convergent. However, the convergence is somewhat slower than was the case when $\omega = \frac{1}{2}$. We recall that the rate of convergence of $\phi_{\text{num}}(\omega;k)$ to $\phi(\omega)$ should be $O(n^{-1})$ for each $0 < \omega < 1$. This corresponds to geometric ($O(2^{-k})$) convergence in k. We can use this observation to accelerate the (numerical) convergence of the data points in Table 6. Below, we work with $x_k = \phi(1 - 2^{-k})$.

If the sequence $x_k \to X$ as $k \to \infty$ with $x_{k+1} - x_k \sim ab^k$ for $|b| < 1$ and $k \to \infty$, then we choose a fixed (large) N, define

$$\tilde{a} = \frac{x_{N+2} - x_{N+1}}{\tilde{b}^{N+1}}, \qquad \tilde{b} = \frac{x_{N+2} - x_{N+1}}{x_{N+1} - x_N}$$



(a) (b)

Figure 1: The functions $\phi_{num}(\omega)$ versus ω for: (a) $k = 5$ and (b) $k = 8$.

and approximate X by

$$X \approx x_N + \frac{\tilde{a}\tilde{b}^N}{1 - \tilde{b}}. \tag{9.5}$$

Applying this to the sequence of ϕ_{num} in Table 6(b) with $N = 6$ and $x_6 = .9057$, $x_7 = .9232$, $x_8 = .9332$ we obtain $\tilde{b} \approx .5714$ and $\tilde{a} \approx .5027$. Then (9.5) leads to $\phi(1) \approx .9465$, and this differs from the theoretical value by only about 0.1%. We note that $O(2^{-k})$ convergence would imply $b = 0.5$. The fact that $\tilde{b}$ somewhat exceeds this value is likely due to the fact that the logarithmic singularity in $\phi'(\omega)$ at $\omega = 1$ causes the convergence to be slightly slower than geometric (say $O(k2^{-k})$) for $\omega \approx 1$.

In Figure 1, we plot ϕ_{num} for $0 < \omega < 1$ (i.e., $0 < n < 2^k$) for $k = 5$ and $k = 8$. As k increases the graphs appear to be converging to some function $\phi(\omega)$, if we renormalized the horizontal axis to the interval $0 < \omega < 1$. Our analytical results predict $\phi(1)$ is finite and $\phi(\omega) - \phi(1) \sim (1 - \omega)\log(1 - \omega)$ as $\omega \to 1^-$; while as $\omega \to 0^+$ we have (roughly) $\phi(\omega) \approx \omega^{.503}$. Thus $\phi'' > 0$ for ω near 1 and $\phi'' < 0$ for ω near 0, with $\phi' \to \infty$ as $\omega \to 0^+$. The graphs in Figure 1 show that $\phi(\omega)$ is convex for most of the range $0 < \omega < 1$, with an abrupt convexity change occurring near $\omega = 0$. The graphs for each of the cases k show a "kink" near $\omega = n2^{-k} = 0$, which indicates a changing convexity. Note however that for (numerically) small values of ω, the convergence of ϕ_{num} to ϕ is extremely slow and thus the graphs cease to be very reliable in this range. If we let $n \to \infty$ and simultaneously $\omega \to 0^+$ then we are exiting the left tail and moving into the central or right tail regimes.

Table 7: Numerical Verification of $\mathbf{E}[\mathcal{H}_n]$

n	$\mathbf{E}(\mathcal{H}_n)$	$\Delta(n)$
10	5.64	–3.946
20	7.74	–3.623
30	9.06	–3.513
50	10.81	–3.409
75	12.25	–3.343
100	13.29	–3.303
200	15.85	–3.220

10 Numerical Studies of the Central Region

We consider the ζ and ν scales, where $0 < L_n^k < 1$ or $1 - L_n^k$ is only algebraically small in n. First we note that the mean height of the binary search tree is given by

$$\mathbf{E}[\mathcal{H}_n] = \sum_{k=1}^{n} k \, \mathsf{Pr}\{\mathcal{H}_n = k\} = \sum_{k=0}^{n} (1 - L_n^k). \tag{10.1}$$

Now define

$$\Delta(n) = \frac{\mathbf{E}[\mathcal{H}_n] - A \log n}{\log\log n} \frac{A-1}{A}, \tag{10.2}$$

where $\mathbf{E}[\mathcal{H}_n]$ is understood to be the exact value of the sum in (10.1).

In Table 7 we give $\Delta(n)$ and $\mathbf{E}[\mathcal{H}_n]$ for n in the range $10 \le n \le 200$. As discussed in Section 2 it was conjectured (cf. (2.9)) that $\Delta(n) \to -1/2$ as $n \to \infty$, while the present conjecture (2.7) has $\Delta(n) \to -3/2$. While Table 7 supports the latter more than the former, even when $n = 200$ we have $\Delta(200) \approx -3.22$. Our analysis suggests that the magnitude of $\Delta(n) - \Delta(\infty)$ is $O((\log\log n)^{-1})$. When $n = 200$, $\log\log n \approx 1.667$ which is not particularly large. To truly test (2.7) and (2.9) we would probably need $\log\log n \approx 10$ and for $n = \exp(e^{10})$ it is not feasible to calculate $\mathbf{E}[\mathcal{H}_n]$ numerically.

If we *assume* that for $n \to \infty$

$$\Delta(n) = \Delta(\infty) + \frac{\bar{c}}{\log\log n}[1 + o(1)], \quad \bar{c} \text{ a constant,}$$

then we can use

$$\Delta(2N) - \Delta(N) \sim \frac{\bar{c}}{\log\log(2N)} - \frac{\bar{c}}{\log\log N}$$

to estimate $\bar{c}$, and then estimate $\Delta(\infty)$ as

$$\Delta(\infty) \approx \Delta(2N) - \frac{\bar{c}}{\log\log(2N)}.$$

Using this idea with $N = 100$ leads to $\bar{c} \approx -1.507$ and then $\Delta(\infty) \approx -2.30$. This suggests that $\Delta(n)$ in Table 7 may converge to a value significantly larger (less negative) than $\Delta(200)$ (≈ -3.22).

Our analysis of the ν-scale involves the unknown function $b(\nu)$. We define

$$b_{\text{num}}(\nu; n) = \sqrt{\log n}\; n^{\nu\log\nu - \nu\log(2e)+1}(1 - L_n^k). \tag{10.3}$$

According to (7.5) we should have, for ν fixed and $n \to \infty$, $b_{\text{num}}(\nu; n) \to b(\nu)$. We have done some numerical experiments with $k = \lfloor 6\log n \rfloor$ (thus $\nu \approx 6$), but found that the sequence (10.3) is not close to converging even when n is as large as 160. It appears that the value of $b(6)$ is of the order 10^{-4}. Indeed our analysis suggests that $b(\nu)$ is very small as $\nu \to \infty$ (cf. (7.8)) and also $b(A) = 0$. Since $b(\nu)$ is only defined for $\nu > A = 4.311\ldots$, it appears that $b(\nu)$ is uniformly small numerically, though asymptotically it is $O(1)$ as $n \to \infty$. This, along with the fact that the asymptotic series on the ν-scale involves inverse powers of $\log n$, makes the numerical computation of $b(\nu)$ difficult. The analysis predicts that $b'(\nu) < 0$ for ν large and $b'(A) > 0$. Thus $b(\nu)$ must have a peak at some $\nu = \nu_0 > A$, but our numerical studies have not been able to confirm this. We note that in order to solve (8.4) numerically, we need the matching condition as $\zeta \to \infty$, and thus the value of $c_1 = b'(A)$.

Alternately, knowing the constant c that appears in the "near left tail" (i.e., the matching region between the ζ and ω scales (cf. (8.30)) is sufficient to uniquely determine $f(\zeta)$ in (8.4). Let us define

$$\begin{aligned} \zeta_1 &= k - A\log n - \frac{1}{2}\frac{A}{A-1}\log\log n \\ \zeta_2 &= k - A\log n - \frac{3}{2}\frac{A}{A-1}\log\log n \end{aligned} \tag{10.4}$$

and note that ζ_1 (resp. ζ_2) corresponds to the conjecture in (2.9) (resp. (2.7)). Define

$$k_*(n) = k_*(n;\lambda) = \lfloor \lambda \log n \rfloor, \quad \text{for} \quad \frac{1}{\log 2} < \lambda < A.$$

Along the sequence $k_*(n)$ we have $\omega \to 0$ and $\zeta_1,\, \zeta_2 \to -\infty$ so we are in the asymptotic matching region of the ω and ζ scales, where (8.30) applies. In Table 8 we take $\lambda = 2$ and consider $L_n^{k_*(n)}$ for $10 \le n \le 200$. We compute ζ_1 and ζ_2 from (10.4) with $k = k_*(n)$, use these values to calculate the right side of (8.30) (up to the constant c) and set the result(s) equal to the (numerical) value of $L_n^{k_*}$. This gives a transcendental equation for c, whose

Table 8:

n	k_*	$L_n^{k_*}$	c (based on ζ_1)	c (based on ζ_2)
10	4	$.698(10^{-1})$	.556	.812
20	5	$.231(10^{-2})$	.601	.989
30	6	$.204(10^{-2})$	.485	.846
40	7	$.615(10^{-2})$	.383	.692
50	7	$.158(10^{-3})$	.448	.832
60	8	$.275(10^{-2})$	.343	.651
70	8	$.223(10^{-3})$	.376	.726
80	8	$.113(10^{-4})$	.408	.798
90	8	$.345(10^{-6})$	.439	.869
100	9	$.150(10^{-3})$	.331	.662
110	9	$.197(10^{-4})$	.348	.702
120	9	$.209(10^{-5})$	.363	.741
130	9	$.177(10^{-6})$	.380	.779
140	9	$.120(10^{-7})$	.395	.816
150	10	$.282(10^{-4})$	.304	.632
160	10	$.619(10^{-5})$	.313	.655
170	10	$.123(10^{-5})$	.322	.677
180	10	$.221(10^{-5})$	.330	.698
190	10	$.360(10^{-7})$	.339	.719
200	10	$.529(10^{-8})$	.347	.740

numerical solution is given in Table 8, using both ζ_1 and ζ_2. While it is plausible that as $n \to \infty$, the numerical values of c do converge to a limit, the oscillatory decrease of $L_n^{k_*}$ as $n \to \infty$ leads to significant oscillations in c. These results support the "double exponential" form of the distribution in the matching region, but they are inconclusive as to the value of the coefficient B of the $\log\log n$ correction term to $\mathbf{E}[\mathcal{H}_n] - A\log n$.

Now consider the differential-delay equation (2.11). Choosing the normalization $F(0) = 1$, we can compute $F(\eta)$ from the iteration scheme

$$F_{N+1}(\eta) = 1 - e^{-1/A}\int_0^{\eta e^{-1/A}} [F_N(u)]^2 du, \qquad F_0(\eta) = 1. \tag{10.5}$$

The Nth iterate $F_N(\eta)$ corresponds to a polynomial approximation of degree $2^N - 1$ to the entire function $F(\eta)$. This method is useful for calculating $F(\eta)$ for moderate values of $|\eta|$. The coefficients in the polynomial approximations have alternating signs, so all terms are positive for $\eta < 0$. For $|\eta| < 3$ we found that the curves $F_7(\eta)$ and $F_8(\eta)$ are virtually indistinguishable. In Table 9, we give $F(\eta)$ to 3 significant digits for certain η in the range [–4,5]. This function grows rapidly as $\eta \to -\infty$ and our numerical studies confirm the asymptotic result in (8.28). Thus (for $F(0) = 1$) we obtain $\nu_2 \approx .066$ in (8.28).

To summarize, our numerical studies confirm the asymptotic analysis on the M, ω, α and j scales. They are not particularly conclusive on the ν and ζ scales, due to the respective asymptotic series involving inverse powers of $\log n$.

Appendix A

We estimate the relative size of the polynomials $P_{2j-2}(n)$, $Q_{2j-4}(n)$ and $R_{2j-6}(n)$ in (6.6). By considering their degrees we obviously have $Q/P = O(n^{-2})$ and $R/P = O(n^{-4})$ for $n \to \infty$ with j fixed. Now let $n, j \to \infty$ at the same rate, with $\alpha = n/j > 1$. In view of (6.16) and (6.18) we have

$$A_m^j = \frac{2^{1-j}}{(j-m-1)!} I(m,j).$$

We have shown that (cf. (6.24))

$$\begin{aligned} \frac{2^n}{n!} P_{2j-2}(n) &\sim \frac{2^{n+1-j} e^{n-j} e^{(j-n)\log j}}{(2\pi j)^{3/2}} \int_0^1 e^{jF(x;\alpha)} G(x;\alpha)dx \\ &\sim \frac{2^{n+1-j}}{2\pi j^2} e^{n-j} e^{(j-n)\log j} \frac{G(x_*)}{\sqrt{|F_{xx}(x_*)|}} e^{jF(x_*)}. \end{aligned} \tag{A.1}$$

Now, $B_m^j = A_m^{j-1} = \frac{2^{2-j}}{(j-m-2)!} I_1$ where

$$I_1 = \frac{1}{2\pi i} \oint z^{-m-1} [\Delta(z)]^{j-m-2} dz$$

and $\Delta(z)$ is defined in Lemma 1. Expanding I_1, by the saddle point method we find that

$$I_1 \sim \frac{1}{\Delta(z_0)} I$$

where I is given by Lemma 1(3). We thus have

$$B_m^j \sim \frac{2(j-m)}{\Delta(z_0)} A_m^j$$

and hence

$$\frac{2^n}{n!} Q_{2j-4}(n) \quad = \quad 2^n \sum_{m=0}^{j-3} B_m^j \frac{1}{(n-2j+m+4)!} \tag{A.2}$$

Table 9: Numerical Evaluation of $F(\eta)$

η	$F(\eta)$
–4	802
–3	44.5
–2.5	15.9
–2	6.98
–1.5	3.64
–1	2.16
–.8	1.80
–.6	1.53
–.4	1.31
–.2	1.14
0	1
.2	.886
.4	.791
.6	.712
.8	.645
1	.588
1.5	.478
2	.399
2.5	.341
3	.296
4	.233
5	.191

$$\sim \quad 2^n \sum_{m=0}^{j-3} \frac{2(j-m)}{\Delta(z_0(m/j))} \frac{1}{n-2j+m+4} \frac{1}{n-2j+m+3} \frac{A_m^j}{(n-2j+m+2)!}$$
$$\sim \quad \frac{2^{n+1-j} e^{n-j} e^{(j-n)\log j}}{(2\pi j)^{3/2}} \int_0^1 \frac{2(1-x)}{j(\alpha+x-2)^2} \frac{1}{\Delta(z_0(x))} e^{jF(x;\alpha)} G(x;\alpha)dx.$$

Evaluating the integral in (A.2) by Laplace's method and comparing the result to (A.1) yields

$$\frac{Q_{2j-4}(n)}{P_{2j-2}(n)} \sim \frac{2}{j} \frac{1-x_*}{\Delta(z_*)} \frac{1}{(\alpha+x_*-2)^2} = \frac{2}{n} \frac{z_*}{1-z_*}. \tag{A.3}$$

A completely analogous argument shows that

$$\frac{R_{2j-6}(n)}{P_{2j-2}(n)} \sim \frac{4}{j^2} \frac{(1-x_*)^2}{\Delta^2(z_*)} \frac{1}{(\alpha+x_*-2)^4} = \frac{4}{n^2} \left(\frac{z_*}{1-z_*} \right)^2. \tag{A.4}$$

On the α scale we have $0 < z_* < 1$ so that the right side of (A.3) is $O(n^{-1})$ and that of (A.4) is $O(n^{-2})$. This shows that $F_n^j/\tilde{F}_n^j = 1 + O(n^{-1})$ for a fixed $\alpha > 1$. Since $c_2 = 0$ in (6.6) we can improve this to $1 + O(n^{-2})$.

For j fixed we can easily show that as $n \to \infty$

$$\frac{Q}{P} \sim \frac{4(j-1)}{n^2}, \qquad \frac{R}{P} \sim \frac{16(j-1)(j-2)}{n^4}.$$

Thus on the j scale $F_n^j/\tilde{F}_n^j = 1 + O(n^{-2})$, and since $c_2 = 0$ this is really $1 + O(n^{-4})$. However, as $\alpha \downarrow 1$ we have

$$1 - z_* \sim \frac{k}{n} \frac{1}{\log n}$$

so that for $\nu = k/\log n$ fixed we have $n(1-z_*) = O(1)$. This shows that P, Q and R (and indeed all the terms in (6.6)) become of comparable magnitude. This observation led us to consider a new scale, namely the ν scale, which we did in Section 7.

Appendix B

We analyze the recurrence (8.16). First we note that as $N \to \infty$, (8.16) resembles the simpler recurrence

$$2c(N) = \sum_{\ell=1}^{N-1} c(\ell)c(N-\ell), \qquad N \geq 2.$$

Taking $c(1) = 1$ and introducing the generating function

$$C(z) = \sum_{N=1}^{\infty} z^N c(N)$$

leads to

$$C^2(z) = 2C(z) - 2z \tag{B.1}$$

so that

$$C(z) = 1 - \sqrt{1 - 2z}$$

and hence

$$c(N) = \binom{2N}{N} \frac{2^{-N}}{2N-1} \sim 2^N N^{-3/2} \frac{1}{2\sqrt{\pi}}, \qquad N \to \infty. \tag{B.2}$$

Now consider (8.16) and let

$$D(z) = \sum_{N=1}^{\infty} d(N) z^N.$$

We thus find that $D(z)$ satisfies

$$D^2(z) = 2D(z) - D\left(\frac{2}{A}z\right) - (A-1)z\frac{d}{dz}D\left(\frac{2}{A}z\right). \tag{B.5}$$

We assume that near the dominant singularity

$$D(z) \sim D(z_*) - \alpha(z_* - z)^{\beta}[1 + O(z_* - z)]; \quad \alpha > 0, \quad 0 < \beta < 1 \tag{B.6}$$

and note that the last two terms in the right side of (B.5) are analytic at z_*. We thus use (B.6) to approximate $D(z)$ and $D^2(z)$ in (B.5) and expand the remaining terms in Taylor series about z_*. We thus obtain

$$\begin{aligned} D^2(z_*) - 2D(z_*) &= -D\left(\frac{2}{A}z_*\right) - z_*(A-1)\frac{d}{dz}D\left(\frac{2}{A}z\right)\Big|_{z=z_*}, \\ -2\alpha D(z_*) &= -2\alpha, \end{aligned} \tag{B.7}$$

and then $2\beta = 1$ and

$$\alpha^2 = A\frac{d}{dz}D\left(\frac{2}{A}z\right)\Big|_{z=z_*} + z_*(A-1)\frac{d^2}{dz^2}D\left(\frac{2}{A}z\right)\Big|_{z=z_*}. \tag{B.8}$$

Hence $D(z_*) = 1$ and then we rewrite (B.7) and (B.8) in terms of $d(N)$, which yields

$$1 = \sum_{N=1}^{\infty} [1 + N(A-1)] \left(\frac{2z_*}{A}\right)^N d(N) \tag{B.9}$$

and

$$\alpha^2 = \frac{1}{z_*} \sum_{N=1}^{\infty} N[1 + N(A-1)] \left(\frac{2z_*}{A}\right)^N d(N). \tag{B.10}$$

We view (B.9) as a transcendental equation for z_*; once it is solved, α is easily computed from (B.10). From (B.6) we find that since $\beta = 1/2$,

$$d(N) \sim z_*^{-N} N^{-3/2} \frac{\alpha\sqrt{z_*}}{2\sqrt{\pi}}, \qquad N \to \infty. \tag{B.11}$$

It follows that the terms in the series (B.9) decay rapidly as $N \to \infty$, as $N^{-1/2}(2/A)^N$. To solve (B.9) we compute the first N_0 of the $d(N)$ numerically from (8.16), then truncate the limits on the sums in (B.9) and (B.10) at $N = N_0$, solve (B.9) for z_* and finally compute α from (B.10). For $10 \leq N_0 \leq 25$ this procedure yields the results below:

N_0	z_*	$1/z_*$	α	$\alpha\sqrt{z_*}/(2\sqrt{\pi})$
10	.25704	3.8903	2.6372	.37717
15	.25702	3.8906	2.6384	.37733
20	.25702	3.8906	2.6384	.37734
25	.25702	3.8906	2.6384	.37734

Thus, this scheme converges rapidly and yields the results in (8.18) and (8.19).

Appendix C

We discuss complex solutions to (8.8) with $\bar{A} = A$. Thus we are interested in solutions to $e^{-z} = 2/(1 + Az)$ in the complex z-plane. We also note that setting $g_L(\zeta) = e^{\zeta/A} G(\zeta)$ in (8.7) leads to

$$G(\zeta + 1) = \frac{2}{A} e^{-1/A} \int_\zeta^\infty G(u)du. \tag{C.1}$$

Since A satisfies $e^{1/A} e^{-1} = 2/A$, it follows that $G'(\zeta) = -e^{-1} G(\zeta - 1)$, which is a retarded differential equation studied in [3]. It admits an infinite number of exponential solutions, whose properties are studied in [3, 17]. These correspond to solutions to (8.7), which we denote by $e^{-a_j \zeta}$ where $a_0 = 1 - 1/A$ and order them as $a_0 < \mathrm{Re}(a_1) < \mathrm{Re}(a_2) < \ldots$. We can take $\mathrm{Im}(a_j) > 0$, $j \geq 1$, since $e^{-\bar{a}_j \zeta}$ is also a solution. The numerical value of a_1 is

$$a_1 = 2.856882062\ldots + i(7.461489285\ldots).$$

Here we shall argue that the solutions other than a_0 cannot be relevant to the present problem, as they lead to solutions of the non-linear problem (8.6) (with $\bar{A} = A$) that corresponds to functions $f(\zeta) = 1 - g(\zeta)$ that are not probability distributions.

First we consider the solution $g_L(\zeta) = k_1 e^{-a_1\zeta}$ to (8.7). By using this as a starting point for solving the non-linear problem (8.6) by successive iterations, we thus obtain

$$g(\zeta) = \sum_{m=1}^{\infty} k_m e^{-ma_1\zeta} \tag{C.2}$$

where the k_m satisfy the recurrence

$$k_m \left[e^{-ma_1} - \frac{2}{ma_1 A + 1} \right] = -\sum_{\ell=1}^{m-1} k_\ell k_{m-\ell} B(1 + Aa_1\ell, 1 + Aa_1(m-\ell)),$$

where B is the Beta function. Setting

$$k_m = k_1^m \frac{[\Gamma(a_1 A + 1)]^m}{\Gamma(ma_1 A + 1)} K_m$$

we obtain the new recurrence

$$K_m \left[2 - (ma_1 A + 1) \left(\frac{2}{1 + a_1 A} \right)^m \right] = \sum_{\ell=1}^{m-1} K_\ell K_{m-\ell}, \qquad K_1 = 1. \tag{C.3}$$

By using ideas completely analogous to those in Appendix B, we find that for $m \to \infty$

$$K_m \sim \mathcal{L}_2 m^{-3/2} \mathcal{L}_1^m$$

where

$$\mathcal{L}_1 \approx 2.017603 - .0570086\ i$$
$$\mathcal{L}_2 \approx .2834684 - .00405513\ i.$$

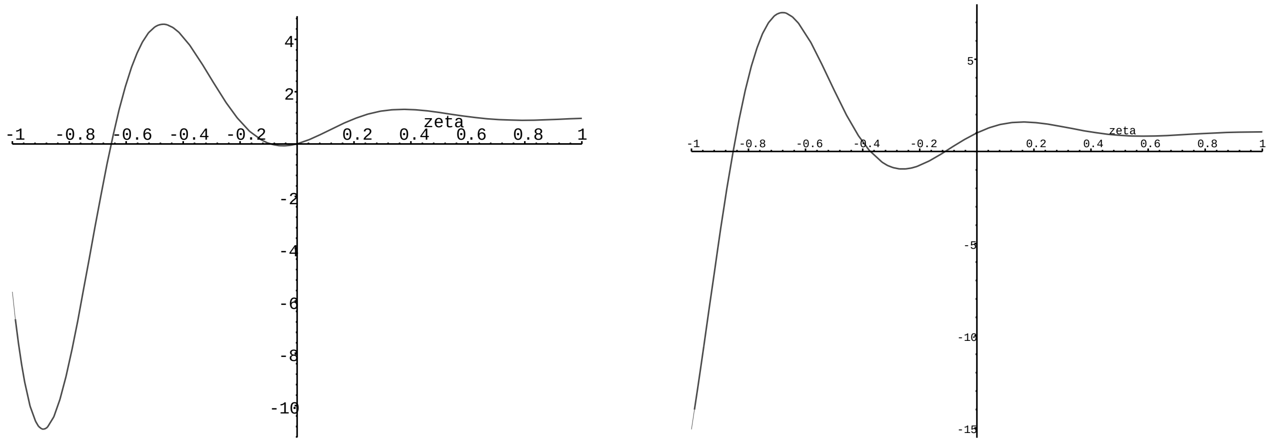
Thus, K_m grows roughly geometrically and (C.2) defines an entire function of ζ. Taking $k_1 = 1$ we plot the real and imaginary parts of $1 - g(\zeta) = f(\zeta)$ in Figure 2. Even over the restricted range $\zeta \in [-1, 1]$, this function(s) oscillates and cannot represent a probability distribution.

The solution $e^{-a_1\zeta}$ can also be excluded by asymptotic matching to the ν-scale expansion. However, a mixture of exponentials, such as

$$g_L(\zeta) = C_0 e^{-a_0\zeta} + C_1 e^{-a_1\zeta} \tag{C.4}$$

cannot be excluded by matching considerations alone. By the translation invariance of (8.6) and the fact that C_0 must be real, we can take $C_0 = 1$. Then (C.4) can be used to construct a solution to the full non-linear problem in the form

$$g(\zeta) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} e^{-\ell a_0\zeta} e^{-ma_1\zeta} k(\ell, m) \tag{C.5}$$



(a) (b)

Figure 2: The functions $1 - \Re(g(\zeta)) = \Re(f(\zeta))$ (left) and $1 - \Im(g(\zeta)) = \Im(f(\zeta))$ (right) for $g_L(\zeta) = e^{-a_1\zeta}$.

where $k(\ell, m)$ are obtained recursively from

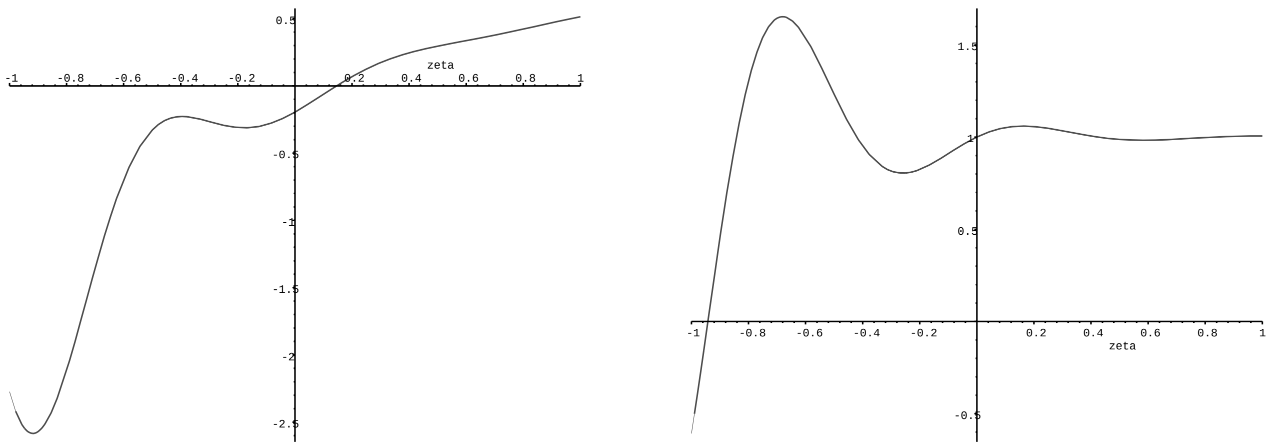
$$\begin{aligned}\left[e^{-\ell a_0} e^{-m a_1} - \frac{2}{1 + A(a_0\ell + a_1 m)}\right] k(\ell, m) &= -\sum_{\ell_1=0}^{\ell} \sum_{m_1=0}^{m} k(\ell_1, m_1) k(\ell - \ell_1, m - m_1) \\ &\times \; B(1 + Aa_0\ell_1 + Aa_1 m_1, 1 + Aa_0(\ell - \ell_1) + Aa_1(m - m_1))\end{aligned}$$

where $k(0,0) = 0$, $k(1,0) = C_0 = 1$ and $k(0,1) = C_1$. Again we plot (cf. Figures 3 and 4) (the real and imaginary parts of) $1 - g(\zeta)$ in (C.5) for various values of C_1. Again we see that the solution is oscillatory. The amplitude of the oscillations is sensitive to the size of C_1, but they are present even for small C_1 (cf. Figure 3).

We have also carried out further numerical experiments that show that solutions to (8.6) involving mixtures of other exponentials (such as $e^{-a_2\zeta}$) lead to more rather than less oscillations. These studies, while not excluding complex solutions with complete vigor, do strongly suggest that they are not at all relevant to the present application.

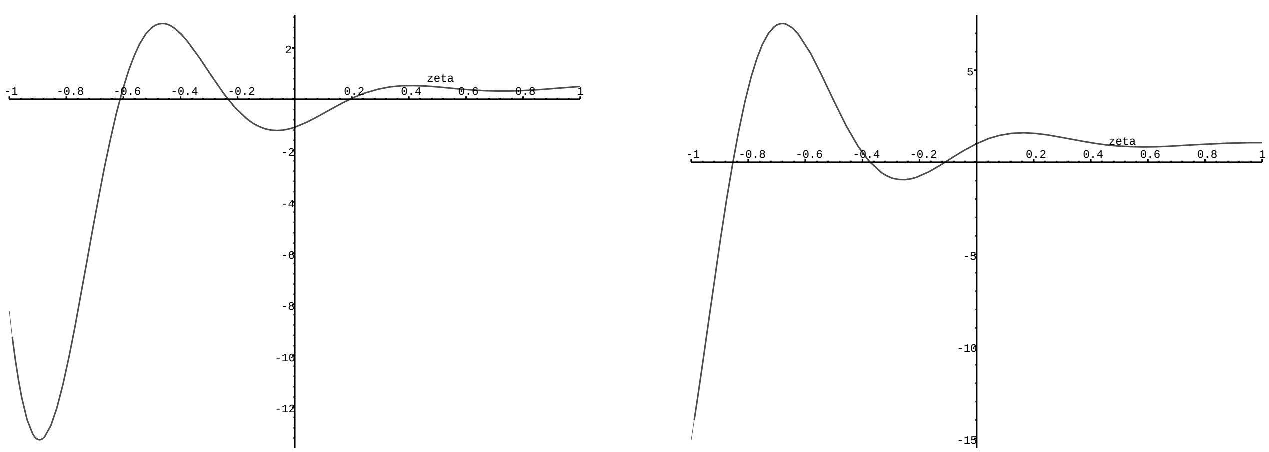
Acknowledgment

We thank two anonymous referees for carefully reading our paper and making many useful suggestions.



(a) (b)

Figure 3: The functions $\Re(f(\zeta))$ (left) and $\Im(f(\zeta))$ (right) for $C_1 = .1$ and $g_L(\zeta) = e^{-a_0\zeta} + C_1 e^{-a_1\zeta}$.



(a) (b)

Figure 4: The functions $\Re(f(\zeta))$ (left) and $\Im(f(\zeta))$ (right) for $C_1 = 1$ and $g_L(\zeta) = e^{-a_0\zeta} + C_1 e^{-a_1\zeta}$.

References

[1] Abramowitz, M. and Stegun, I., eds. *Handbook of Mathematical Functions*. Wiley, New York, 1962.

[2] G. Andrews, R. Askey and R. Roy, *Special Functions*, Cambridge University Press, 1999.

[3] R. Bellman and K. Cooke, Differential Difference Equations, Academic Press, 1963.

[4] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1978.

[5] L. Devroye, A Note on the Height of Binary Search Trees, *J. of the ACM*, 33, 489-498, 1986.

[6] L. Devroye, Branching Processes in the Analysis of the Height of Trees, *Acta Informatica*, 24, 277-298, 1987.

[7] L. Devroye and B. Reed, On the Variance of the Height of Random Binary Search Trees, *SIAM. J. Computing*, 24, 1157–1162, 1995.

[8] M. Drmota, An Analytic Approach to the Height of Binary Search Trees, *Algorithmica*, 29, 89–119, 2001.

[9] M. Drmota, An Analytic Approach to the Height of Binary Search Trees. II, preprint.

[10] M. V. Fedoryuk, *Asymptotic Analysis*, Springer Verlag, Berlin, 1993.

[11] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM J. Disc. Methods*, 3, 216-240, 1990.

[12] E. L. Ince, *Ordinary Differential Equations*, Dover, New York 1956.

[13] D. Knuth, *The Art of Computer Programming. Sorting and Searching*, Second Edition, Addison-Wesley, 1998.

[14] C. Knessl and W. Szpankowski, Quicksort Algorithm Again Revisited, *Discrete Mathematics and Theoretical Computer Science*, 3, 43-64, 1999.

[15] C. Knessl and W. Szpankowski, Limit Laws for Heights in Generalized Tries and PATRICIA Tries, *Proc. LATIN'2000*, Punta del Este, Uruguay, Lecture Notes in Computer Science, No. 1776, 298-307, 2000.

[16] C. Knessl and W. Szpankowski, Asymptotic Behavior of the Height in a Digital Search Tree and the Longest Phrase of the Lempel-Ziv Scheme, *SIAM J. Compt.*, 30, 923-964, 2000.

[17] C. Lange and R. Miura, Singular Perturbation Analysis of Boundary Value Problems for Differential-Difference Equations.V.Small Shifts with Layer Behavior, SIAM J. Appl. Math., 54, 1994, pp. 249–272.

[18] G. Louchard, Exact and Asymptotic Distributions in Digital and Binary Search Trees, *RAIRO Theoretical Inform. Applications*, 21, 479-495, 1987.

[19] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.

[20] J. Murdock, *Perturbations: Theory and Methods*, SIAM, Philadelphia, 1999.

[21] A. H. Nayfeh, *Introduction to Perturbation Techniques*, John Wiley & Sons, New York, 1981.

[22] A. Odlyzko, Asymptotic Enumeration, in *Handbook of Combinatorics*, Vol. II, (Eds. R. Graham, M. Grötschel and L. Lovász), Elsevier Science, 1063-1229, 1995.

[23] B. Pittel, On Growing Random Binary Trees, *J. Math. Anal. Appl.*, 103, 461-480, 1984.

[24] B. Reed, How Tall is a Tree, *Proc. STOC*, 479-483, Portland, 2000.

[25] J. M. Robson, The Height of Binary Search Trees, *Austral. Comput. J.*, 11, 151-153, 1979.

[26] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, John Wiley & Sons, New York, 2001.

[27] R. Wong, *Asymptotic Approximations of Integrals*, Academic Press, Boston, 1989.