LARGE DEVIATIONS FOR THE LEAVES IN SOME RANDOM TREES

WLODEK BRYC,* ** AND

DAVID MINDA,* *** *University of Cincinnati*

SUNDER SETHURAMAN,**** *Iowa State University*

Abstract

Large deviation principles and related results are given for a class of Markov chains associated to the 'leaves' in random recursive trees and preferential attachment random graphs, as well as the 'cherries' in Yule trees. In particular, the method of proof, combining analytic and Dupuis–Ellis-type path arguments, allows for an explicit computation of the large deviation pressure.

Keywords: Large deviation; central limit; preferential attachment; planar oriented; uniformly random trees; leaves; cherries; Yule; random Stirling permutations

2000 Mathematics Subject Classification: Primary 60F10
Secondary 05C80

1. Introduction and results

In this paper we consider large deviations and related laws of large numbers and central limit theorems for a class of Markov chains associated to the number of leaves, or nodes of degree one, in preferential attachment random graphs and random recursive trees, and also the number of 'cherries', or pairs of leaves with a common parent, in Yule trees. These random graphs model various networks such as pyramid schemes, chemical polymerization, the Internet, social structures, genealogical families, among others. In particular, the leaf and cherry counts in these models are of interest, and have concrete interpretations.

Define the nondecreasing Markov chain $\{Z_n : n \geq 1\}$ starting from the initial state $Z_1 = k_0 \geq 0$ by its one-step transitions

$$\Pr(Z_{n+1} - Z_n = v \mid Z_n) = \begin{cases} 1 - \dfrac{Z_n}{s_n} & \text{if } v = 1, \\ \dfrac{Z_n}{s_n} & \text{if } v = 0, \end{cases} \tag{1.1}$$

where $\{s_n : n \geq 1\}$ is a sequence of positive numbers such that

$$s_n \geq k_0 + n - 1 \quad \text{and} \quad \frac{s_n}{n} \to \alpha \quad \text{for some } 1 < \alpha < \infty \tag{1.2}$$

with convention that $0/0 = 0$. Additionally, we also consider two special sequences:

$$s_n = n, \qquad \alpha = 1, \quad \text{and} \quad k_0 = 0, 1, \tag{1.3}$$

and

$$s_n = \frac{n}{2}, \qquad \alpha = \tfrac{1}{2}, \quad \text{and} \quad k_0 = 0. \tag{1.4}$$

The Markov chain Z_n, with respect to certain s_ns and αs, will be seen to represent the count of leaves in preferential attachment and recursive trees, and the count of cherries in Yule trees. For most of these models, a law of large numbers (LLN) and a central limit theorem (CLT) with respect to Z_n have been proved.

Then, characterizing the associated large deviations is a natural problem which gives insight into the properties of rare events, and seems less studied in random graphs. Previous large deviations work on related models has concentrated on analytic methods with respect to some urn schemes, not applicable in our setting [25], or on extensions of the Dupuis–Ellis weak convergence approach (cf. [20]) to allocation counts different than ours [21], [44]. We also note that some exponential bounds via martingale concentration inequalities are found in the case where s_n is linear with slope α [15]. See also [4], [6], [12], [18], [19], and [28] for other types of large deviations work in various random tree models.

Our main results are to prove a large deviation principle (LDP) for Z_n/n with an explicitly computed 'pressure', or Legendre transform of the associated rate function (Theorem 1.1). Such explicit computations are not commonplace, and our ordinary differential equation (ODE) method, as described below, is quite different from the methods in [25], where a quasi-linear partial differential equation (PDE) is solved, or in [44], where a finite-dimensional minimization problem is obtained.

In addition, aside from a LLN, which is trivially obtained, we prove a CLT for Z_n through complex variable arguments with the pressure (Theorem 1.3). These proofs of the LLN and CLT, although indirect, serve as alternate arguments when the LLN and CLT are already known in model contexts.

In Subsection 1.2 we more carefully define the random graph models considered, provide the related literature, and discuss applications of our results through Propositions 1.1–1.4. We remark that Proposition 1.2 gives a quenched LLN, CLT, and LDPs for the count of certain leaves in a randomized preferential attachment model which involves a random s_n.

The large deviation arguments to handle the chain when $\alpha > 1$, that is, under assumption (1.2), and the cases $\alpha = 1$ and $\alpha = \tfrac{1}{2}$, that is, under assumptions (1.3) and (1.4), respectively, make use of two different methods: an ODE method under assumption (1.2), and a singularity analysis approach under assumptions (1.3) and (1.4). The ODE approach relies on a large deviation principle for the path interpolation of $Z_{\lfloor nt \rfloor}/n$ (Theorem 1.2), perhaps of interest in itself, that we establish by the Dupuis–Ellis weak convergence approach.

Our ODE technique to prove the LDP for Z_n/n is to consider the recurrence relation for $m_n(\lambda) = \mathrm{E}[\exp\{\lambda Z_n\}]$ obtained from (1.1):

$$m_{n+1}(\lambda) = (1 - \mathrm{e}^\lambda)\frac{m_n'(\lambda)}{s_n} + \mathrm{e}^\lambda m_n(\lambda). \tag{1.5}$$

Dividing through by $m_n(\lambda)$, we write

$$\frac{m_{n+1}(\lambda)}{m_n(\lambda)} = \frac{1 - \mathrm{e}^\lambda}{s_n/n}\frac{m_n'(\lambda)}{nm_n(\lambda)} + \mathrm{e}^\lambda. \tag{1.6}$$

The idea now is to take the limit on n in the above display. When the 'pressure' Λ exists, it satisfies $\Lambda(\lambda) = \lim_{n\to\infty}(1/n)\log m_n(\lambda)$. In this case, it is natural to suppose that the limits

$$\Lambda'(\lambda) = \lim_{n\to\infty} \frac{m_n'(\lambda)}{nm_n(\lambda)}, \tag{1.7}$$

$$e^{\Lambda(\lambda)} = \lim_{n\to\infty} \frac{m_{n+1}(\lambda)}{m_n(\lambda)} \tag{1.8}$$

both exist. Then, from (1.6) we can write the ODE

$$e^{\Lambda(\lambda)} = \frac{1 - e^\lambda}{\alpha}\Lambda'(\lambda) + e^\lambda, \qquad \Lambda(0) = 0. \tag{1.9}$$

This equation has unique solution (1.13), below.

The main task is to show that the pressure and limits (1.7) and (1.8) exist. But, the pressure exists as a consequence of the path LDP for $Z_{\lfloor nt \rfloor}/n$ by the contraction principle. We note, in principle, that we can try to compute the pressure or the rate function from (1.14), below, by the calculus of variations, but we found it difficult to solve the associated Euler equations for (5.1), below.

Finally, we show that (1.7) and (1.8) exist by extending $m_n(\lambda)$ to the complex plane, and then analyzing its zeros and analytic properties. These estimates are also useful for the CLT arguments.

The second approach to prove the LDP under assumptions (1.3) and (1.4), when $\alpha = 1$ and $\alpha = \frac{1}{2}$, uses the fact that s_n is linear with slope α. In this approach, in the spirit of [25], we compute the pressure from analysis of singularities for the generating function $G(\lambda, z) = \sum_{n\geq 1} m_n(\lambda)z^{n-1}$. From (1.5) we can write the linear PDE

$$\frac{\partial G}{\partial z}(1 - e^\lambda z) + \frac{e^\lambda - 1}{\alpha}\frac{\partial G}{\partial \lambda} = e^\lambda G. \tag{1.10}$$

We can solve this PDE implicitly, and locate, at least heuristically, a singular point. Then, formally, from root test asymptotics, the pressure would be the reciprocal of the location of the singularity.

The difficulty is in establishing the analyticity of the solution and identifying its singularity. Flajolet *et al.* [25] used this program to obtain large deviations and the CLT for a certain class of urn models. However, the cases where s_n is linear with slope $\alpha = \frac{1}{2}$, 1, 2 and, more generally, the urns associated with noninteger s_n are not covered by their arguments. On the other hand, we are able to supply the needed analyticity and singularity identification when s_n has slopes $\alpha = \frac{1}{2}$, 1, 2, and in this way prove the LDP for Z_n/n (Theorem 1.1) in these cases.

The plan of the paper is to state the results in Subsection 1.1, discuss applications to random graphs in Subsection 1.2, prove the path LDP (Theorem 1.2) in Section 2, prove the LDP for Z_n/n (Theorem 1.1) and associated LLN and CLT (Theorem 1.3) in Section 3, reprove Theorem 1.1 for $s_n = 2n$ by singularity analysis in Section 4, and conclude in Section 5.

1.1. Results

We recall the setting for large deviations. A sequence $\{X_n\}$ of random variables with values in a separable complete metric space $\mathcal{X}$ satisfies the LDP with speed n and rate function $I\colon \mathcal{X} \to [0, \infty]$ if I has compact level sets $\{x\colon I(x) \leq a\}$ for $a \geq 0$, and for every Borel

set $U \in \mathcal{B}_{\mathcal{X}}$,

$$- \inf_{x \in U^\circ} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \Pr(X_n \in U)$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \log \Pr(X_n \in U)$$

$$\leq - \inf_{x \in \bar{U}} I(x).$$

(Here U° is the interior of U and $\bar{U}$ is the closure of U.)

Often the rate function is given in terms of the Legendre transform of the pressure $\Lambda(\cdot)$ when it exists. When $\mathcal{X} = \mathbb{R}$, this representation takes the form

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \log \Lambda(\lambda)\}, \tag{1.11}$$

where we recall that

$$\Lambda(\lambda) := \lim_{n \to \infty} \frac{1}{n} \log \mathrm{E}[\exp\{\lambda Z_n\}]. \tag{1.12}$$

Now recall the Markov chain Z_n corresponding to sequence $\{s_n\}$, (1.1).

Theorem 1.1. *Suppose that one of the conditions (1.2), (1.3), or (1.4) holds. Then, the sequence Z_n/n satisfies the LDP with speed n and strictly convex rate function I given by (1.11) with pressure*

$$\Lambda(\lambda) = -\log\left(\frac{\alpha}{e^\lambda - 1} \int_0^\lambda \left(\frac{e^s - 1}{e^\lambda - 1}\right)^{\alpha - 1} ds\right) \quad for \ \lambda \neq 0 \tag{1.13}$$

and $\Lambda(0) = 0$.

Remark 1.1. For $\alpha = \frac{1}{2}$ or for integer $\alpha \geq 1$, the integral in (1.13) can be evaluated explicitly; cf. (3.4), (3.5), and (4.1), below.

We now consider the LDP for the family of stochastic processes $\{X_n(t) \colon 0 \leq t \leq 1\}$ obtained by linear interpolation of the Markov chain (1.1),

$$X_n(t) := \frac{1}{n} Z_{\lfloor nt \rfloor - k_0 + 1} + \frac{nt - \lfloor nt \rfloor}{n} (Z_{\lfloor nt \rfloor - k_0 + 2} - Z_{\lfloor nt \rfloor - k_0 + 1}) \quad \text{for } t \geq \frac{k_0}{n}$$

and $X_n(t) := t$ for $0 \leq t \leq k_0/n$. The trajectories of $X_n(t)$ are nondecreasing Lipschitz functions with constant at most 1.

Theorem 1.2. *Suppose that condition (1.2) holds. As a sequence of $C([0,1]; \mathbb{R})$-valued random variables, X_n satisfies the LDP with speed n and convex rate function $I \colon C([0,1]; \mathbb{R}) \to [0, \infty]$ given by*

$$I(\varphi) = \int_0^1 \left(\dot{\varphi}(t) \log \frac{\alpha t \dot{\varphi}(t)}{\alpha t - \varphi(t)} + (1 - \dot{\varphi}(t)) \log \frac{\alpha t (1 - \dot{\varphi}(t))}{\varphi(t)}\right) dt \tag{1.14}$$

if $\varphi(0) = 0$, $\varphi(t)$ is differentiable, and $0 \leq \dot{\varphi} \leq 1$ for almost all t, and the integral converges; otherwise, $I(\varphi) = \infty$.

By the contraction principle, Theorem 1.2 implies the LDP for Z_n/n with rate function given by the variational expression

$$I(x) = \inf\{I(\varphi) \colon \varphi(0) = 0, \ \varphi(1) = x\}. \tag{1.15}$$
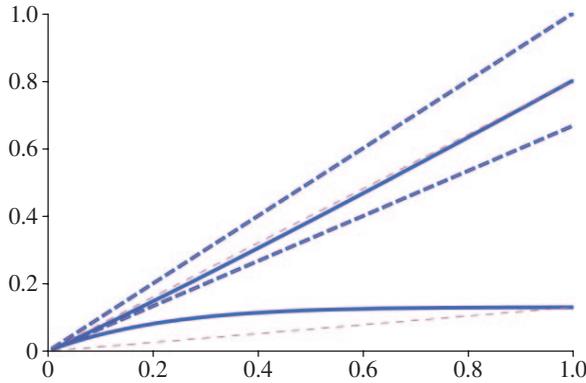
FIGURE 1: Thick curves are numerical solutions of the Euler equations for (1.15) with $\alpha = 2$ for $x = 0.13, \frac{2}{3}, 0.85, 1$. Dashed, thin or thick, lines are straight lines from $(0, 0)$ to $(1, x)$.

In general, optimal trajectories are not straight lines—exceptions are the LLN trajectory $\varphi_\alpha(t) = t\alpha/(\alpha + 1)$ and the extreme case $\varphi(t) = t$. But, the optimal trajectories try to stay near the LLN line (for which $I(\varphi_\alpha) = 0$) to minimize cost before going to destination x (cf. Figure 1).

Lemmas for the proof of Theorem 1.1 give normal approximation. The law of large numbers also follows from Theorem 1.1.

Theorem 1.3. *Suppose that one of the conditions (1.2), (1.3), or (1.4) holds. Then, we have*

$$\frac{Z_n}{n} \to \frac{\alpha}{\alpha + 1} \quad \text{almost surely (a.s.)}$$

and also

$$\frac{1}{\sqrt{n}}(Z_n - \mathrm{E}[Z_n]) \xrightarrow{\mathrm{D}} N(0, \sigma^2), \quad \text{where} \quad \sigma^2 = \frac{\alpha^2}{(1 + \alpha)^2(2 + \alpha)},$$

and '$\xrightarrow{\mathrm{D}}$' denotes convergence in distribution.

1.2. Applications to random graph models

As alluded to in the introduction, the Markov chain Z_n, depending on parameters, represents the number of leaves in at least two random graph models, that is, preferential attachment graphs with linear-type weights, and uniformly and planar oriented trees. Also, Z_n in a particular case corresponds to the count of cherries in Yule trees.

1.2.1. *Preferential attachment graphs.* Preferential attachment graphs have a long history dating back to Yule (cf. [35]). However, since the work of Barabasi–Albert [1], [5], these graphs have been of recent interest with respect to modeling of various 'real-world' networks such as the Internet (WWW), and social and biological communities. Leaves, or nodes with degree one, in these networks of course represent sites with one link, or members at the periphery. (See books [11], [15], and [23] for more discussion.)

The idea is to start with an initially connected graph G_1 with a finite number of vertices, and say no self-loops (so that the vertices have well-defined degrees). At step 1, add another vertex, and connect it to a vertex x of G_1 preferentially, that is, with probability proportional to its weight, $f(d_x)/\sum_{y \in G_1} f(d_y)$, to form a new graph G_2. Continue in this way by adding a new

vertex and connecting it preferentially to form G_k for $k \geq 1$. Here, the 'weight' of a vertex is a function f of its degree d_x. When $f : \mathbb{N} \to \mathbb{R}_+$ is increasing, already well-connected vertices tend to become better connected, a sort of reinforcing effect. We note when the initial graph G_1 is a tree, the later graphs $\{G_n\}$ are also trees.

Our results will be applicable to linear weights, $f(k) = k + \beta$ for $\beta > -1$, which correspond to certain power-law mean degree sequences. Namely, let $T_n(k)$ be the number of vertices in G_n with degree k for $k \geq 1$. It was shown, by martingale arguments in [9] and [36], and by embedding into branching processes in [40], that the LLN $T_n(k)/n \to r_k$ a.s. holds, where

$$r_k = \frac{2 + \beta}{k + \beta} \prod_{j=1}^{k} \frac{j + \beta}{j + 2 + 2\beta} = O\left(\frac{1}{k^{3+\beta}}\right).$$

We note that a corresponding CLT for $T_n(k)$ is proved in [36]. Part of the appeal, with respect to applications, is that the parameter β can sometimes be matched to empirical network data where similar power-law behavior is observed.

Let d_{G_1} and $|G_1|$ be the total degree and number of vertices in G_1, respectively. Define also $k_0^{G_1}$ as the number of leaves in G_1. For the number of vertices with degree 1, or the leaves $T_n = T_n(1)$, we have the following result which in part reproves the associated LLN and CLT.

Proposition 1.1. *The count T_n is the Markov chain Z_n with*

$$s_n = \frac{1}{1 + \beta}(d_{G_1} + 2(n - 1) + (|G_1| + n - 1)\beta),$$

$k_0 = k_0^{G_1}$, *and* $\alpha = (2 + \beta)/(1 + \beta)$. *Hence, as* $\alpha > 1$, *the LLN, CLT in Theorem 1.3, and LDPs in Theorems 1.1 and 1.2 apply.*

Proof. The count T_n increases by one in the next step when a nonleaf is selected, and remains the same when a leaf is chosen. Since at each step the total degree of the graph augments by two, at step n, the total degree of G_n is $d_{G_1} + 2(n - 1)$, and so the total weight of G_n with $|G_1| + n - 1$ vertices is $d_{G_1} + 2(n - 1) + (|G_1| + n - 1)\beta$. Therefore, the probability at step n that a vertex $x \in G_n$ is selected is $(d_x + \beta)/(d_{G_1} + 2(n - 1) + (|G_1| + n - 1)\beta)$. Then, at step n, given that the leaves have total weight $T_n + T_n\beta$, the probability that a leaf is selected is $T_n(1 + \beta)/(d_{G_1} + 2(n - 1) + (|G_1| + n - 1)\beta)$. Therefore, T_n can be identified with the Markov chain Z_n with $s_n = (d_{G_1} + 2(n - 1) + (|G_1| + n - 1)\beta)/(1 + \beta)$, $k_0 = k_0^{G_1}$, and $\alpha = (2 + \beta)/(1 + \beta)$. The condition $s_n \geq k_0 + n - 1$ is straightforwardly verified.

We can also randomize the model by adding a random number of edges at each step. Let $\{\gamma_i\}$ be a sequence of independent, identically distributed random variables on $\mathbb{N}$ with finite mean $\bar{\gamma} = \mathrm{E}[\gamma_1] < \infty$. Consider the following evolution given a realization of the sequence $\{\gamma_i\}$. At step $n \geq 1$, we add a new vertex to the graph G_n and connect it to a node selected preferentially from G_n with $\gamma_n + 1$ directed edges put between them, that is, one edge directed to the new vertex and the remaining γ_n directed towards the selected node in G_n. Here, to select preferentially means that a node in $x \in G_n$ is selected with probability proportional to weight $d_x^{\mathrm{in}} + \beta$, where d_x^{in} is the in-degree of x. For simplicity, to allow the full range $\beta > -1$, in the following we will assume that all nodes in the initial graph G_1 have in-degree at least 1. The effect of these random edge additions with respect to $\{\gamma_i\}$ is to randomize further the weight given the nodes in the graph. The deterministic model, that is, the model given above, is when $\mathrm{P}(\gamma_1 = 1) \equiv 1$ and the directions on edges are not taken in account. We note that the randomized model is

similar to the one in [3]. See also [16] for a more general randomized model, and also [19] and [32] for other random edge schemes.

In our randomized model we now define the notion of generalized leaves, that is, those nodes which connect to exactly one other vertex, or in other words, those vertices with in-degree equal to 1. Let $T_n^{\text{gen}} = T_n^{\text{gen}}(1)$ denote the number of generalized leaves at step n, and let $d_{G_1}^{\text{in}}$ be the total in-degree of G_1. Define also k_0^{gen} as the number of generalized leaves in G_1. The following result gives new quenched LLN, CLT, and LDPs with respect to T_n^{gen}.

Proposition 1.2. *Given the sequence $\{\gamma_i\}$, T_n^{gen} is the Markov chain Z_n with $k_0 = k_0^{\text{gen}}$, $s_1 = (d_{G_1}^{\text{in}} + |G_1|\beta)/(1 + \beta)$ and, for $n \geq 2$,*

$$s_n = \frac{1}{1 + \beta}\left(d_{G_1}^{\text{in}} + \sum_{i=1}^{n-1}(\gamma_i + 1) + (|G_1| + n - 1)\beta\right).$$

Hence, a.s. with respect to $\{\gamma_i\}$, since $s_n/n \to \alpha = (\bar{\gamma} + 1 + \beta)/(1 + \beta) > 1$, the LLN, CLT in Theorem 1.3, and LDPs in Theorems 1.1 and 1.2 hold for $\{Z_n\}$ conditioned on $\{\gamma_i\}$.

Proof. Similar to the leaves in the deterministic model, at each step, the generalized leaf count increases by one when the new vertex connects to a nongeneralized leaf, and remains the same when the new vertex links to a generalized leaf. In step $n \geq 1$, the total in-degree of the graph increases by $\gamma_n + 1$. Also, the total weight of the graph at step 1 is $d_{G_1}^{\text{in}} + |G_1|\beta$, and at step $n \geq 2$ is $d_{G_1}^{\text{in}} + \sum_{i=1}^{n-1}(\gamma_i + 1) + (|G_1| + n - 1)\beta$. Then, the probability that the new vertex links to vertex x at step 1 is $(d_x^{\text{in}} + \beta)/(d_{G_1}^{\text{in}} + |G_1|\beta)$ and at step $n \geq 2$ is $(d_x^{\text{in}} + \beta)/(d_{G_1}^{\text{in}} + \sum_{i=1}^{n-1}(\gamma_i + 1) + (|G_1| + n - 1)\beta)$. Therefore, analogous to the deterministic model, the probability a leaf is selected at step 1 is $T_n^{\text{gen}}(1 + \beta)/(d_{G_1}^{\text{in}} + |G_1|\beta)$ and at step $n \geq 2$ is $T_n^{\text{gen}}(1 + \beta)/(d_{G_1}^{\text{in}} + \sum_{i=1}^{n-1}(\gamma_i + 1) + (|G_1| + n - 1)\beta)$. Hence, T_n^{gen} is seen to be the Markov chain Z_n with s_n and k_0 as desired, and satisfying $s_n \geq k_0 + n - 1$.

1.2.2. *Random recursive trees.*

Random recursive trees are also well-established models, dating to the 1960s, with applications to data sorting and searching, pyramid schemes, spread of epidemics, chemical polymerization, family trees (stemma) of copies of ancient manuscripts, etc. Leaves in these trees correspond to 'shutouts' with respect to pyramid schemes, nodes with small 'affinity' in polymerization models, 'terminal copies' in stemma of manuscripts, etc. See [41], [42], and the references therein (e.g. [37]) for more discussion. Following the proof of Proposition 1.3, below, we also mention connections with Stirling permutations.

These recursive schemes form a sequence of trees. We start with a single vertex labeled 0 with degree 1 (e.g. connected to a node outside the system), and then attach a new vertex at step $n \geq 1$, labeled n, to one of the n nodes already present. When the choice is made uniformly over the labels $0, 1, \ldots, n - 1$, the model forms a growing *uniformly* recursive tree. However, when the vertex, say $x \in \{0, 1, \ldots, n - 1\}$, is chosen with probability proportional to its degree d_x, and the new vertex is inserted at random uniformly in one of the d_x gaps between its $d_x - 1$ children (the left and right of all child labels joining x are also considered gaps), a *plane oriented* tree is grown. Here, unlike for the uniformly recursive tree scheme, different orders of labels at each distance from the root 0 give rise to distinct trees.

Let $R_n^{\text{unif}}(k)$ and $R_n^{\text{plan}}(k)$ be the numbers of vertices at step n with degree k in the uniform and planar oriented schemes, respectively. With respect to both types of recursive schemes, LLN and CLTs for $R_n(k)$ have also been proved by combinatorial, urn, and martingale methods (see [23, Chapter 4], [29], and [41]). Part of the next result reproves the LLN and CLT with respect to the leaves $R_n^{\text{unif}} = R_n^{\text{unif}}(1)$ and $R_n^{\text{plan}} = R_n^{\text{plan}}(1)$.

Proposition 1.3. *The count R_n^{unif} is the Markov chain Z_n with $k_0 = 1$, $s_n = n$, and $\alpha = 1$. However, R_n^{plan} is the chain Z_n with $k_0 = 1$, $s_n = 2n - 1$, and $\alpha = 2$. Hence, with respect to R_n^{unif} and R_n^{plan}, the LLN, CLT, and LDP in Theorems 1.1 and 1.3 apply. In addition, with respect to R_n^{plan}, the path LDP in Theorem 1.2 holds.*

Proof. The counts R_n^{unif} and R_n^{plan} increase by one in the next step when a nonleaf is selected, and remain the same when a leaf is chosen. With respect to the uniform scheme, the probability that at step n a vertex $x \in \{0, \ldots, n-1\}$ is selected is $1/n$. Hence, the probability that a leaf is selected at this step is R_n^{unif}/n. Since $R_1^{\text{unif}} = 1$ (the 0th labeled node), it follows that R_n^{unif} is identified with the Markov chain Z_n with $k_0 = 1$, $s_n = n$, and $\alpha = 1$.

On the other hand, with respect to the planar oriented scheme (which is similar to preferential attachment with $\beta = 0$), at each step the total degree of the graph increases by two, and so the total degree of the tree at step n, noting that the degree of 0 is initially 1, is $2(n-1)+1 = 2n-1$. Therefore, the probability that at step n a vertex $x \in \{0, \ldots, n-1\}$ is selected is $d_x/(2n-1)$. Correspondingly, at step n, as there are R_n^{plan} leaves each with degree 1, the probability that a leaf is selected is $R_n^{\text{plan}}/(2n-1)$. Since initially $R_1^{\text{plan}} = 1$, R_n^{plan} is seen to be the chain Z_n with $k_0 = 1$, $s_n = 2n - 1$, and $\alpha = 2$.

We now comment on recent connections of planar oriented trees with Stirling permutations (cf. [30] and [31]). A Stirling permutation of length $2n$ is a permutation of the multiset $\{1, 1, 2, 2, \ldots, n, n\}$ such that, for each $i \le n$, the elements occurring between the two is are larger than i (cf. [27]).

It turns out that each permutation is a distinct code for a plane-oriented recursive tree with $n + 1$ vertices. Indeed, quoting from [30], consider the *depth first* walk which starts at the root and goes first to the leftmost daughter of the root, explores that branch (recursively, using the same rules), returns to the root, and continues to the next daughter, and so on. Each edge is passed twice in the walk, once in each direction. Label the edges in the tree according to the order in which they were added—edge j is added at step j and connects vertex j to a previously labeled vertex. The plane recursive tree is coded by the sequence of labels passed by the depth first walk. With respect to a tree with $n + 1$ vertices, the code is of length $2n$, where each of the labels $1, 2, \ldots, n$ appears twice. Adding a new vertex means inserting the pair $(n + 1)(n + 1)$ in the code in one of the $2n + 1$ places.

In a Stirling permutation $a_1 a_2 \cdots a_{2n}$, the index $1 \le i \le 2n$ is a plateau if $a_i = a_{i+1}$ (where $a_{2n+1} = 0$). Janson [30] showed that the number of leaves in a plane-oriented tree with $n + 1$ vertices is the number of plateaux in a random Stirling permutation of length $2n$. See [30] for more details.

1.2.3. *Yule trees.* Since Yule's influential 1924 paper [43], Yule trees, among other processes, have been used widely to model phylogenetic evolutionary relationships between species (see [2] for an interesting essay). In particular, the counts of various shapes and features of these trees can be studied, and matched to empirical data to test evolutionary hypotheses. In [34], an LLN and CLT is proved for the number of cherries, or pairs of leaves with a common parent, in Yule trees. Associated confidence intervals are computed, and some empirical data sets are examined to see their compatibility with 'Yule tree' genealogies. Other related statistical tests and limit results can be found in [7], [8], [26], and [39].

In the Yule tree process, we start with a root vertex. It will split into two daughter nodes at step 1, each of which is equally likely to split into two children at step 2. At step n, one of the n leaves in the tree is chosen at random, and it then splits into two daughters, and so on. Let

C_n be the number of cherries at step $n \geq 1$. The following proposition reproves the LLN and CLT for the cherry counts C_n, and also states an associated LDP.

Proposition 1.4. *The cherry count C_n is the Markov chain Z_n with $k_0 = 0$, $s_n = n/2$, and $\alpha = \frac{1}{2}$. Hence, the LLN, CLT, and LDP in Theorems 1.1 and 1.3 hold.*

Proof. The counts C_n increase by one in the next step when a leaf not part of a cherry is selected, and remains the same when a leaf in a cherry pair is chosen. At step n, as one of the n leaves is selected at random, the probability that a leaf in a cherry pair is taken is $2C_n/n$. As initially $C_1 = 0$ (only the root node is present), C_n is identified with the chain Z_n with $k_0 = 0$, $s_n = n/2$, and $\alpha = \frac{1}{2}$.

2. Proof of Theorem 1.2

We follow the method and notation of Dupuis and Ellis [20]. Although some arguments are analogous to those found in [20, Chapter 6], which considers random walk models with time-homogeneous continuous statistics, and [44], where a different model with a time singularity at $t = 0$ is examined, for completeness we give all details as several differ, especially in the lower bound proof.

Let $X_j^n = Z_{j-k_0+1}/n$ for $k_0 \leq j \leq n$, and set $X_j^n = j/n$ for $0 \leq j \leq k_0$. Then, noting (1.1), given X_j^n, we have $X_{j+1}^n = X_j^n + v_j^n/n$, where v_j^n has Bernoulli distribution $\rho_{\sigma_n(j/n), X_j^n}$. Here,

$$\sigma_n(t) = \begin{cases} s_{\lfloor nt-k_0+1 \rfloor}/n & \text{for } t \geq k_0/n, \\ 0 & \text{for } t < k_0/n, \end{cases}$$

$$\rho_{\sigma,x}(A) = \frac{x}{\sigma}\delta_0(A) + \left(1 - \frac{x}{\sigma}\right)\delta_1(A) \quad \text{for } A \subset \mathbb{R}, \ 0 \leq x \leq \sigma,$$

and $\rho_{0,0} := \delta_1$.

Define $X_\cdot^n$ as the polygonal interpolated path connecting points $(j/n, X_j^n)$ for $0 \leq j \leq n$. Also, for probability measures $\mu \ll \nu$ such that $\mathrm{d}\mu = f \, \mathrm{d}\nu$, define $R(\mu \| \nu) = \int f \log f \, \mathrm{d}\nu$, the relative entropy; set $R(\mu \| \nu) = \infty$ when μ is not absolutely continuous with respect to ν.

Let $h \colon C([0,1], \mathbb{R}) \to \mathbb{R}$ be a bounded continuous function. To prove Theorem 1.2, we need only establish Laplace principle upper and lower bounds [20, p. 74]. The upper bounds are to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{E}[\exp\{-nh(X_\cdot^n)\}] \leq - \inf_{\varphi \in C([0,1],\mathbb{R})} \{I(\varphi) + h(\varphi)\}$$

for a rate function I. The lower bounds are to prove the reverse inequality:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{E}[\exp\{-nh(X_\cdot^n)\}] \geq - \inf_{\varphi \in C([0,1],\mathbb{R})} \{I(\varphi) + h(\varphi)\}.$$

Define, for $k_0 + 1 \leq j \leq n$, noting that $X_j^n = j/n$ for $j \leq k_0$ is deterministic,

$$W^n(j, x_j, \ldots, x_{k_0+1}) = -\frac{1}{n} \log \mathrm{E}[\exp\{-nh(X_\cdot^n)\} \mid X_j^n = x_j, \ldots, X_{k_0+1}^n = x_{k_0+1}]$$

and

$$W^n := W^n(k_0, \varnothing) = -\frac{1}{n} \log \mathrm{E}[\exp\{-nh(X_\cdot^n)\}].$$

Then, by the Markov property, for $k_0 + 1 \leq j \leq n - 1$,

$$\exp\{-nW^n(j, x_j, \ldots, x_{k_0+1})\}$$
$$= \mathrm{E}[\exp\{-nW^n(j+1, X_{j+1}^n, x_j, \ldots, x_{k_0+1})\} \mid X_j^n = x_j, \ldots, X_{k_0+1}^n = x_{k_0+1}]$$
$$= \int \exp\left\{-nW^n\left(j+1, x_j + \frac{v}{n}, x_j, \ldots, x_{k_0+1}\right)\right\} \rho_{\sigma_n(j/n), x_j}(\mathrm{d}v).$$

By a property of relative entropy [20, Proposition 1.4.2(a)], for $k_0 + 1 \leq j \leq n - 1$,

$$W^n(j, x_j, \ldots, x_{k_0+1})$$
$$= -\frac{1}{n}\log \int \exp\left\{-nW^n\left(j+1, x_j + \frac{y}{n}, x_j, \ldots, x_{k_0+1}\right)\right\} \rho_{\sigma_n(j/n), x_j}(\mathrm{d}v)$$
$$= \inf_{\gamma}\left\{\frac{1}{n}R(\gamma \| \rho_{\sigma_n(j/n), x_j}) + \int W^n\left(j+1, x_j + \frac{y}{n}, x_j, \ldots, x_1\right)\gamma(\mathrm{d}y)\right\}.$$

Also, the boundary condition $W^n(n, x_n, \ldots, x_{k_0+1}) = h(x.)$ holds with respect to the linearly interpolated path $x. = x.^n$ connecting $\{(\ell/n, x_\ell/n)\}_{k_0 \leq \ell \leq n}$.

The basic observation in the Dupuis–Ellis method is that $W^n(j, x_j, \ldots, x_{k_0+1})$ satisfies a control problem (see [20, Section 3.2]) whose solution, for $k_0 \leq j \leq n - 1$, is

$$V^n(j, x_j, \ldots, x_{k_0+1}) = \inf_{\{v_i^n\}} \bar{\mathrm{E}}_{j, x_j, \ldots, x_{k_0+1}}\left[\frac{1}{n}\sum_{i=j}^{n-1} R(v_i^n(\cdot) \| \rho_{\sigma_n(i/n), \bar{X}_i^n}) + h(\bar{X}_.^n)\right].$$

Here, $v_i^n(\mathrm{d}y) = v_i^n(\mathrm{d}y; x_{k_0}, \ldots, x_i)$ is a Bernoulli distribution given $x_{k_0}, \ldots, x_i$ for $k_0 \leq i \leq n - 1$ and in the display $v_i^n(\cdot) = v_i^n(\mathrm{d}y \mid \bar{X}_{k_0}^n, \ldots, \bar{X}_i^n)$; $\{\bar{X}_i^n; 0 \leq i \leq n\}$ is the adapted path satisfying $\bar{X}_l^n = l/n$ for $0 \leq l \leq k_0$ and $\bar{X}_{l+1}^n = \bar{X}_l^n + \bar{Y}_l^n/n$ for $k_0 \leq l \leq n - 1$, where $\bar{Y}_l^n$, conditional on $(\bar{X}_l^n, \ldots, \bar{X}_{k_0}^n)$, has distribution $v_l^n(\cdot)$; $\bar{X}_.^n$ is the interpolated path with respect to $\{\bar{X}_l^n\}$; and $\bar{\mathrm{E}}_{j, x_{j_2}, \ldots, x_{k_0+1}}$ denotes the conditional expectation with respect to the $\bar{X}_.^n$ process given the values $\{\bar{X}_l^n = x_l : k_0 + 1 \leq l \leq j\}$ at step $k_0 + 1 \leq j \leq n$. The boundary conditions are $V^n(n, x_n, \ldots, x_{k_0+1}) = h(x.)$ and

$$V^n(k_0, \varnothing) = V^n = \inf_{\{v_j^n\}} \bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=k_0}^{n-1} R(v_j^n(\cdot) \| \rho_{\sigma_n(j/n), \bar{X}_j^n}) + h(\bar{X}_.^n)\right]. \tag{2.1}$$

In particular, by [20, Corollary 5.2.1],

$$W^n = -\frac{1}{n}\log \mathrm{E}[\exp\{-nh(X_.^n)\}] = V^n. \tag{2.2}$$

The goal will be to take Laplace limits using this representation. To simplify later expressions, we will take $v_j^n = \delta_1$ for $0 \leq j \leq k_0 - 1$ when $k_0 \geq 1$.

2.1. Upper bound

To establish the upper bound, we first put the controls $\{v_j^n\}$ into continuous-time paths. Let $v^n(\mathrm{d}y \mid t) = v_j^n(\mathrm{d}y)$ for $t \in (j/n, (j+1)/n]$, $0 \leq j \leq n - 1$, and $v^n(\mathrm{d}y \mid 0) = v_0^n$. Define

$$v^n(A \times B) = \int_B v^n(A \mid t)\,\mathrm{d}t$$

for $A \subset \mathbb{R}$ and $B \subset [0, 1]$. Also, define the piecewise constant path $\tilde{X}^n(t) = \bar{X}^n_j$ for $t \in (j/n, (j+1)/n]$, $0 \le j \le n-1$, and $\tilde{X}^n(0) = 0$. Then,

$$V^n = \inf_{\{v^n_j\}} \bar{\mathbb{E}}\left[\int_0^1 R(v^n(\cdot \mid t) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)}) \, dt + h(\bar{X}^n_\cdot)\right].$$

From this control representation, as $|V^n| = |W^n| \le \|h\|_\infty$ and $\rho_{\sigma, x}$ is supported on $K = \{0, 1\}$, for each n, there are $\{v^n_j\}$ supported on K and corresponding $v^n(dy \times dt) = v^n(dy \mid t) \times dt$ such that

$$W^n + \epsilon = V^n + \epsilon \ge \bar{\mathbb{E}}\left[\int_0^1 R(v^n(\cdot \mid t) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)}) \, dt + h(\bar{X}^n_\cdot)\right].$$

As the sets K and

$$\Gamma = \{\varphi \in C([0, 1]; \mathbb{R}) : \varphi(0) = 0, \varphi \uparrow, \text{ Lipschitz, with bound 1}\}$$

are compact on $\mathbb{R}$ and $C([0, 1], \mathbb{R})$, respectively, and $\{v^n_j\}$ are probability measures on K and $\{\bar{X}^n_\cdot\} \subset \Gamma$, by Prokhorov's theorem, the distributions of $(v^n(dy \times dt), \bar{X}^n_\cdot)$ have a subsequence which converges weakly to a limit distribution governing $(v, \bar{X}_\cdot)$, where, for each realization, v is a probability measure on $K \times [0, 1]$ and $\bar{X}_\cdot \in \Gamma$. More precisely, let $(\bar{\Omega}, \bar{F}, \bar{P})$ be a probability space where $v \colon \bar{\Omega} \to$ probability measures on $K \times [0, 1]$, and $\bar{X} \colon \bar{\Omega} \to \Gamma$. Then, [20, Lemma 3.3.1] shows that v is the subsequential weak limit of v^n, and, $\bar{P}$-a.s., for $\omega \in \bar{\Omega}$,

$$v(A \times B \mid \omega) = \int_B v(A \mid t, \omega) \, dt$$

for some kernel $v(dy \mid t, \omega)$ on K given t and ω.

Now, by the same proof given for [20, Theorem 5.3.5] (only [20, Equation (5.12)] in the theorem statement differs; in our context μ there is replaced by $\rho_{\sigma_n(j/n), \bar{X}^n_j}$), as K is compact, $(v^n, \bar{X}^n, \tilde{X}^n)$ has a subsequential limit in distribution to $(v, \bar{X}, \bar{X})$ (the last coordinate with respect to $D([0, 1], \mathbb{R})$). Also, $\bar{P}$-a.s., for all $t \in [0, 1]$,

$$\bar{X}(t) = \int_{\mathbb{R} \times [0,1]} y \, v(dy \times ds) = \int_0^t \int_K y \, v(dy \mid s) \, ds.$$

In particular, $\bar{P}$-a.s., $\dot{\bar{X}}(t) = \int_K y \, v(dy \mid t)$.

By the Skorokhod representation theorem we can assume that $(v^n, \bar{X}^n, \tilde{X}^n) \to (v, \bar{X}, \bar{X})$ converges a.s. In particular, $\bar{X}^n$ converges uniformly to $\bar{X}$ a.s., and it is clear that $\tilde{X}^n$ converges uniformly to $\bar{X}$ a.s. as $\bar{X}$ is continuous (see [20, p. 154]).

Let $\mu_\delta(\cdot) = (1 - \delta)^{-1} \mathbf{1}_{[\delta, 1]}(\cdot)$, and define measures $v^n_\delta(dy \times dt) := v^n(dy \mid t) \times \mu_\delta \, dt$. Then, by [20, Lemma 1.4.3(f)] we have, for $0 < \delta < 1$,

$$\int_\delta^1 R(v^n(\cdot \mid t) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)}) \, dt = (1 - \delta) R(v^n_\delta(dy \times dt) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)} \times \mu_\delta \, dt).$$

Write

$$\liminf_{n \to \infty} V^n \ge \liminf_{n \to \infty} \bar{\mathbb{E}}\left[\int_\delta^1 R(v^n(\cdot \mid t) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)}) \, dt + h(\bar{X}^n_\cdot)\right]$$

$$\ge \liminf_{n \to \infty} \bar{\mathbb{E}}[(1 - \delta) R(v^n_\delta(dy \times dt) \| \rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)} \times \mu_\delta \, dt) + h(\bar{X}_\cdot)]$$

$$\ge \bar{\mathbb{E}}\left[\int_\delta^1 R(v(\cdot \mid t) \| \rho_{\alpha t, \bar{X}(t)}) \, dt + h(\bar{X}_\cdot)\right],$$

where in the last line we used Fatou's lemma, noting lower semi-continuity of R, $v_\delta^n(\mathrm{d}y \times \mathrm{d}t) \to v(\mathrm{d}y \mid t) \times \mu_\delta \, \mathrm{d}t$ a.s., and $\rho_{\sigma_n(\lfloor nt \rfloor/n), \tilde{X}^n(t)}$ converges in distribution to $\rho_{\alpha t, \bar{X}(t)}$ for $t \in [\delta, 1]$ a.s. since $\sigma_n(\lfloor nt \rfloor/n) \to \alpha t$, $\tilde{X}^n(t) \to \bar{X}(t)$ uniformly on $[0, 1]$ a.s., and $\rho_{\sigma, x}$ is continuous on $\{(\sigma, x) \colon \delta \le \sigma \le 1, \ 0 \le x \le \sigma\}$ (cf. [20, Section 6.2]).

By [20, Lemma 3.3.3(b)],

$$R(v(\cdot \mid t) \| \rho_{\alpha t, \bar{X}(t)}) \ge L\left(\alpha t, \bar{X}(t), \int y v(\mathrm{d}y \mid t)\right),$$

where

$$L(t, x, y) = \sup_\theta \left\{ \theta y - \log \int \mathrm{e}^{\theta z} \rho_{t,x}(\mathrm{d}z) \right\}. \tag{2.3}$$

We note that, for $t > 0$, $L(t, x, y)$ diverges when $x = 0$ and $0 \le y < 1$, and $x = t$ and $0 < y \le 1$, but is finite otherwise, and in this case evaluates to

$$L(t, x, y) = y \log\left(\frac{ty}{t - x}\right) + (1 - y) \log\left(\frac{t(1 - y)}{x}\right), \tag{2.4}$$

understood with the usual convention that $0 \log(0) = 0$.

Since $\int y v(\mathrm{d}y \mid t) = \dot{\bar{X}}(t)$, we have

$$\liminf_{n \to \infty} V^n \ge \bar{\mathrm{E}}\left[\int_\delta^1 L(\alpha t, \bar{X}(t), \dot{\bar{X}}(t)) \, \mathrm{d}t + h(\bar{X}_\cdot)\right].$$

As $L \ge 0$ and $\bar{X}(\cdot) \in \Gamma$, letting $\delta \downarrow 0$, we obtain

$$\liminf_{n \to \infty} V^n \ge \inf_{\varphi \in \Gamma} \int_0^1 L(\alpha t, \varphi(t), \dot{\varphi}(t)) \, \mathrm{d}t + h(\varphi).$$

Taking into account (2.2), the upper bound holds with $I(\varphi) = \int_0^1 L(\alpha t, \varphi(t), \dot{\varphi}(t)) \, \mathrm{d}t$ when $\phi \in \Gamma$, and $I(\varphi) = \infty$ otherwise.

We observe that I is convex as, for fixed $t > 0$, $(0, t] \times [0, 1] \ni (x, y) \mapsto L(\alpha t, x, y)$ is convex by [20, Lemma 1.4.3(b)]. Finally, I also has compact level sets by the proof of [20, Proposition 6.2.4]. (We only note that we use $0 \le \dot{\varphi} \le 1$ instead of their Condition 6.2.1, so that the uniform integrability on line 21 of [20, p. 161] holds, and for the last display on p. 163 of [20], for fixed $t > 0$ and $0 \le y \le 1$, that $[0, 1] \ni x \mapsto L(\alpha t, x, y)$ is lower semi-continuous.)

2.2. Lower bound

In the following, for typographical convenience, we write $\mathrm{E}[X; \mathbb{A}]$ for $\int_\mathbb{A} X \, \mathrm{d} \mathrm{P}$. Now let $\psi^* \in \Gamma$ be such that $I(\psi^*) < \infty$, and fix a bounded, continuous (in the sup norm) function $h \colon C([0, 1]; \mathbb{R}) \to \mathbb{R}$. In view of (2.2), we need only show, for each $\varepsilon > 0$, that

$$\limsup_{n \to \infty} V^n \le I(\psi^*) + h(\psi^*) + 8\varepsilon. \tag{2.5}$$

Step 1. Our first goal is to replace ψ^* by its appropriate regularization. We use the trick of considering a convex combination of paths,

$$\psi_\theta(t) = (1 - \theta)\psi^*(t) + \theta t \quad \text{for } 0 \le \theta \le 1.$$

Since $\|\psi_\theta - \psi^*\|_\infty \leq 2\theta$, it is clear, for small enough $\theta > 0$, that $|h(\psi_\theta) - h(\psi^*)| < \varepsilon$. Furthermore, since I is finite on the line $\ell(t) = t$ with slope 1, by convexity of I (cf. the discussion at the end of Subsection 2.1), for small enough $\theta > 0$,

$$I(\psi_\theta) \leq (1-\theta)I(\psi^*) + \theta I(\ell) \leq I(\psi^*) + \varepsilon.$$

We therefore fix $\theta > 0$ such that $I(\psi_\theta) < I(\psi^*) + \varepsilon$ and $h(\psi_\theta) < h(\psi^*) + \varepsilon$.

Next, following [20, p. 82], we write

$$\varphi_\kappa(t) = \int_0^t \gamma_\kappa(s)\, ds, \tag{2.6}$$

where

$$\gamma_\kappa(t) = \kappa \int_{i/\kappa}^{i+1/\kappa} \dot{\psi}_\theta(s)\, ds$$

for $t \in (i/\kappa, (i+1)/\kappa]$, $0 \leq i \leq \kappa - 1$, and $\gamma_\kappa(0) = \gamma_\kappa(1/\kappa)$. For large enough κ, we have

$$\frac{\varphi_\kappa(t)}{t} \geq \theta > 0, \tag{2.7}$$

$$h(\varphi_\kappa) \leq h(\psi^*) + 2\varepsilon, \tag{2.8}$$

$$I(\varphi_\kappa) \leq I(\psi^*) + 2\varepsilon. \tag{2.9}$$

Inequality (2.7) is a property of ψ_θ, and is preserved by (2.6). Since

$$\lim_{\kappa \to \infty} \sup_{t \in [0,1]} |\psi_\theta(t) - \varphi_\kappa(t)| = 0,$$

inequality (2.8) follows from continuity of h by our choice of θ. Then, since

$$\dot{\varphi}_\kappa(t) = \kappa \int_{\lfloor t\kappa \rfloor/\kappa}^{\lceil t\kappa \rceil/\kappa} \dot{\psi}_\theta(s)\, ds \to \dot{\psi}_\theta(t) \quad \text{as } \kappa \uparrow \infty,$$

and $(t, x, y) \mapsto L(\alpha t, x, y)$ (cf. (2.3) and (2.4)) for $\alpha > 1$ is bounded and uniformly continuous on the compact set

$$\{(t, x, y): 0 \leq t \leq 1,\ \theta t \leq x \leq t,\ \theta \leq y \leq 1\},$$

by the dominated convergence theorem, $\lim_{\kappa \to \infty} I(\varphi_\kappa) = I(\psi_\theta)$. Inequality (2.9) follows due to our choice of θ.

We now fix κ such that the above bounds hold.

Since (2.9) implies that $I(\varphi_\kappa) < \infty$, we now choose a $0 < \delta < \frac{1}{3}$ such that

$$\int_0^\delta L(\alpha t, \varphi_\kappa(t), \dot{\varphi}_\kappa(t))\, dt < \varepsilon. \tag{2.10}$$

We will also need an estimate on s_n. From assumptions (1.2), there exist $0 < \eta < \frac{1}{2}$ and $k_1 \geq k_0 + 1$ such that, for $n \geq k_1$,

$$\eta \leq \frac{n}{s_{n-k_0+1}} \leq 1 - \eta. \tag{2.11}$$

With respect to η and θ, we additionally impose the condition that δ satisfies

$$-\delta \log(\delta\eta\theta) < \frac{\varepsilon}{10}. \tag{2.12}$$

Step 2. We now build a sequence of controls based on φ_κ. Recall that we already set $v_j^n(\mathrm{d}y) = \delta_1$ for $0 \leq j \leq k_0 - 1$ when $k_0 \geq 1$. Define

$$v_j^n(\mathrm{d}y; x_{k_0}, \ldots, x_j) = \begin{cases} \rho_{\sigma_n(j/n), x_j} & \text{for } k_0 \leq j \leq k_1, \\ \rho_{1, 1-\dot{\varphi}_\kappa(j/n)} & \text{for } j \geq k_1 + 1. \end{cases}$$

Note that, for $j \geq k_1 + 1$, v_j^n does not depend on auxiliary inputs $x_{k_0}, \ldots, x_j$, and is in fact the Bernoulli distribution with success probability $\dot{\varphi}_\kappa(j/n)$.

Define also $\bar{X}_l^n = l/n$ for $0 \leq l \leq k_0$ and $\bar{X}_{j+1}^n = \bar{X}_j^n + \bar{Y}_j^n/n$ for $j \geq k_0$, where

$$\bar{P}(\bar{Y}_j^n \in \mathrm{d}y \mid \bar{X}_0^n, \ldots, \bar{X}_j^n) = v_j(\mathrm{d}y; \bar{X}_0^n, \ldots, \bar{X}_j^n).$$

Thus, for $j \geq 1$, $\bar{X}_j^n = (1/n)\sum_{\ell=0}^{j-1} Y_\ell^n$, where the $\{Y_j^n\}_{j \geq k_1+1}$ are independent Bernoulli random variables with corresponding means $\{\dot{\varphi}_\kappa(j/n)\}_{j \geq k_1+1}$.

Step 3. We now collect some useful estimates.

(A) Since $\bar{Y}_0^n \equiv 1$ (when $k_0 = 0$, recall that $\rho_{\sigma,0} = \delta_1$), and the increments are at most one, we have $1/n \leq \bar{X}_j^n \leq j/n$ for $1 \leq j \leq n$.

(B) We have

$$\lim_{n \uparrow \infty} \sup_{0 \leq j/n \leq 1} \left| \bar{X}_j^n - \frac{1}{n}\sum_{l=0}^j \dot{\varphi}_\kappa\left(\frac{l}{n}\right) \right| = 0 \quad \text{a.s.} \tag{2.13}$$

Indeed, for large enough n, as $0 \leq \dot{\varphi}_\kappa \leq 1$, $(1/n)\sum_{j=0}^{k_1} |\dot{\varphi}_\kappa(j/n)| + k_1/n < n^{-1/8}$. Then, by Doob's maximal inequality,

$$\begin{aligned}
\bar{P}&\left(\sup_{0 \leq j/n \leq 1} \left| \bar{X}_j^n - \frac{1}{n}\sum_{l=0}^j \dot{\varphi}_\kappa(l/n) \right| > \frac{\theta}{2n^{1/8}} \right) \\
&\leq Cn^{1/2}\bar{E}|\bar{X}_n^n - \bar{E}[X_n^n]|^4 \\
&\leq \frac{1}{n^{7/2}}\left[C\sum_{l=0}^n \dot{\varphi}_\kappa\left(\frac{l}{n}\right)\left(1 - \dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right)\left(1 - 3\dot{\varphi}_\kappa\left(\frac{l}{n}\right) + 3\dot{\varphi}_\kappa^2\left(\frac{l}{n}\right)\right) \right. \\
&\qquad\qquad \left. + C\left(\sum_{l=0}^n \dot{\varphi}_\kappa\left(\frac{l}{n}\right)\left(1 - \dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right)\right)^2 \right] \\
&\leq Cn^{-3/2}, \tag{2.14}
\end{aligned}$$

where C is a constant changing from line to line.

(C) For $0 \leq j \leq n$, from (2.7), it follows that

$$\frac{1}{n}\sum_{l=0}^j \dot{\varphi}_\kappa\left(\frac{l}{n}\right) \geq \theta\frac{j}{n}. \tag{2.15}$$

(D) Let $j \geq k_1 + 1$. Noting that $1/n \leq \bar{X}_j^n \leq j/n$ (cf. part (A)) and bounds (2.11), we have

$$\eta \leq 1 - \frac{j}{s_{j-k_0+1}} \leq 1 - \frac{\bar{X}_j^n}{\sigma_n(j/n)} \leq 1, \qquad \frac{1}{s_{j-k_0+1}} \leq \frac{\bar{X}_j^n}{\sigma_n(j/n)} \leq 1 - \eta.$$

Hence, $L(\sigma_n(j/n), \bar{X}_j^n, \dot{\varphi}_\kappa(j/n))$ can be well evaluated (cf. (2.3) and (2.4)), and we may rewrite the relative entropy as

$$R(v_j^n \| \rho_{\sigma_n(j/n), \bar{X}_j^n}) = \dot{\varphi}_\kappa\left(\frac{j}{n}\right) \log\left(\frac{\dot{\varphi}_\kappa(j/n)}{1 - \bar{X}_j^n/\sigma_n(j/n)}\right) + \left(1 - \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right) \log\left(\frac{1 - \dot{\varphi}_\kappa(j/n)}{\bar{X}_j^n/\sigma_n(j/n)}\right)$$

$$= L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right).$$

Furthermore, as $0 \leq \dot{\varphi}_\kappa \leq 1$, we have

$$L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right)$$

$$= \left[\dot{\varphi}_\kappa\left(\frac{j}{n}\right) \log \dot{\varphi}_\kappa\left(\frac{j}{n}\right) + \left(1 - \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right) \log\left(1 - \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right)\right]$$

$$- \dot{\varphi}_\kappa\left(\frac{j}{n}\right) \log\left(1 - \frac{\bar{X}_j^n}{\sigma_n(j/n)}\right) - \left(1 - \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right) \log\left(\frac{\bar{X}_j^n}{\sigma_n(j/n)}\right)$$

$$\leq 0 - \log \eta + \log(s_{j-k_0+1})$$

$$\leq \log\left(\frac{j}{\eta^2}\right), \tag{2.16}$$

where, for the last inequality, we have used (2.11) again.

Step 4. We now argue (2.5) via representation (2.1). Let

$$\mathbb{A} = \left\{\sup_{0 \leq j \leq n-1} \left|\bar{X}_j^n - \frac{1}{n} \sum_{l=0}^{j} \dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| > \frac{\theta}{2n^{1/8}}\right\}.$$

Since $R(v_j^n \| \rho_{\sigma_n(j/n), \bar{X}_j^n}) = 0$ for $k_0 \leq j \leq k_1$, the sum in (2.1) equals

$$\bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=k_0}^{n-1} R(v_j^n \| \rho_{\sigma_n(j/n), \bar{X}_j^n})\right] = \bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=k_1+1}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right)\right]$$

$$= \bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=k_1+1}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}\right]$$

$$+ \bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=k_1+1}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}^c\right]$$

$$=: A_1 + A_2. \tag{2.17}$$

Step 5. We now treat term A_1 in (2.17). Combining (2.16) with (2.14), we obtain, for large n,

$$\bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=k_1+1}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \sup_{0\leq j\leq n-1}\left|\bar{X}_j^n - \frac{1}{n}\sum_{l=0}^{j}\dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| > \frac{\theta}{2n^{1/8}}\right]$$

$$\leq \log\left(\frac{n}{\eta^2}\right)\bar{\mathrm{P}}\left[\sup_{0\leq j/n\leq 1}\left|\bar{X}_j^n - \frac{1}{n}\sum_{l=0}^{j}\dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| > \frac{\theta}{2n^{1/8}}\right]$$

$$\leq C\log\left(\frac{n}{\eta^2}\right)n^{-3/2}$$

$$< \varepsilon. \tag{2.18}$$

Step 6. For term A_2 in (2.17), we split it into two sums depending on whether index $j \leq \delta n$ or $j \geq \delta n$ (recall δ from (2.10)):

$$\bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=k_1+1}^{\lfloor\delta n\rfloor} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}^c\right]$$

$$+ \bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=\lceil\delta n\rceil}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}^c\right]$$

$$=: B_1 + B_2. \tag{2.19}$$

Step 7. To estimate B_1, we further divide it into two terms corresponding to sums on indices $j \leq n^{7/8}$ and $n^{7/8} \leq j \leq \delta n$:

$$\bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=k_1+1}^{\lfloor n^{7/8}\rfloor} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}^c\right]$$

$$+ \bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=\lceil n^{7/8}\rceil}^{\lfloor\delta n\rfloor} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right); \mathbb{A}^c\right]$$

$$=: D_1 + D_2.$$

Term D_1, using (2.16), is bounded for all large n by

$$n^{-1/8}\log\left(\frac{n^{7/8}}{\eta^2}\right) < \varepsilon. \tag{2.20}$$

For term D_2, using the second-line equality stated in (2.16), we note that

$$\left|L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right)\right| \leq -\log\eta + \log\sigma_n\left(\frac{j}{n}\right) - \log\bar{X}_j^n$$

$$= -\log\eta + \log\left(\frac{s_{j-k_0+1}}{n}\right) - \log\bar{X}_j^n.$$

Then, D_2 is bounded in absolute value by

$$-\delta \log \eta + \frac{1}{n} \sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \left| \log\left(\frac{s_{j-k_0+1}}{n}\right)\right|$$

$$-\bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \log \bar{X}_j^n;\ \sup_{0\le j\le n-1}\left|\bar{X}_j^n - \frac{1}{n}\sum_{l=0}^{j}\dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| \le \frac{\theta}{2n^{1/8}}\right].$$

Now, note that (2.11) implies that, for $j \ge k_1 + 1$,

$$\frac{j}{(1-\eta)n} \le \frac{s_{j-k_0+1}}{n} \le \frac{j}{\eta n}.$$

Then, as $0 < \delta < \frac{1}{3}$ and $0 < \eta < \frac{1}{2}$, we have, for large n,

$$\frac{1}{n} \sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \left|\log\left(\frac{s_{j-k_0+1}}{n}\right)\right| \le 2\int_0^\delta \max\left\{\left|\log\left(\frac{x}{1-\eta}\right)\right|, \left|\log\left(\frac{x}{\eta}\right)\right|\right\}\mathrm{d}x$$

$$\le -4\delta \log \delta - 2\delta \log \eta.$$

Also, noting (2.15), we have, for large n,

$$-\bar{\mathrm{E}}\left[\frac{1}{n} \sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \log \bar{X}_j^n;\ \sup_{0\le j\le n-1}\left|\bar{X}_j^n - \frac{1}{n}\sum_{l=0}^{j}\dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| \le \frac{\theta}{2n^{1/8}}\right]$$

$$\le -\frac{1}{n}\sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \log\left(\frac{\theta j}{n} - \frac{\theta}{2n^{1/8}}\right)$$

$$\le -\frac{1}{n}\sum_{j=\lceil n^{7/8}\rceil}^{\lfloor \delta n \rfloor} \log\left(\frac{\theta j}{2n}\right)$$

$$\le -2\int_0^\delta \log\frac{\theta x}{2}\,\mathrm{d}x$$

$$\le -6\delta \log \delta - 2\delta \log \theta.$$

By combining these estimates, D_2 is bounded by a function of δ, η, and θ, which, given (2.12), can be made small:

$$D_2 \le -10\delta \log \delta - 3\delta \log \eta - 2\delta \log \theta < \varepsilon. \tag{2.21}$$

Step 8. We now estimate term B_2 in (2.19). Note, for $n > \delta^{-8}$, by (2.15), the event

$$\left\{\sup_{0\le j\le n-1}\left|\bar{X}_j^n - \frac{1}{n}\sum_{l=0}^{j}\dot{\varphi}_\kappa\left(\frac{l}{n}\right)\right| \le \frac{\theta}{2n^{1/8}}\right\} \subset \left\{\inf_{j\ge \delta n}\bar{X}_j^n \ge \frac{\delta\theta}{2}\right\}.$$

Hence, for large n,

$$B_2 \le \bar{\mathrm{E}}\left[\frac{1}{n}\sum_{j=\lceil \delta n \rceil}^{n-1} L\left(\sigma_n\left(\frac{j}{n}\right), \bar{X}_j^n, \dot{\varphi}_\kappa\left(\frac{j}{n}\right)\right);\ \mathbb{A}^{\mathrm{c}} \cap \left\{\inf_{j\ge \delta n}\bar{X}_j^n \ge \frac{\delta\theta}{2}\right\}\right].$$

Now, $L(t, x, y)$ is continuous, and, therefore, also bounded and uniformly continuous on the compact set (cf. (2.3) and (2.4))

$$\left\{ (t, x, y) \colon \frac{\delta}{1 - \eta} \leq t \leq \frac{1}{\eta}, \; \frac{\delta \theta}{4} \leq x \leq \left(1 - \frac{\eta}{2} \right) t, \; 0 \leq y \leq 1 \right\}. \tag{2.22}$$

Also, from (2.11), when $\delta \leq j/n \leq 1$, we have $\delta/(1 - \eta) \leq \sigma_n(j/n) \leq 1/\eta$ and both $\bar{X}_j^n$ and $(1/n) \sum_{l=0}^{j} \dot{\varphi}_\kappa(l/n) \leq j/n = \sigma_n(j/n)(j/s_{j-k_0+1}) \leq \sigma_n(j/n)(1 - \eta)$ for all large n. In addition, on the set $\{\inf_{j \geq \delta n} \bar{X}_j^n \geq \delta \theta/2\}$, and by (2.15), when $\delta n \leq j \leq n$, we have both $\bar{X}_j^n$ and $(1/n) \sum_{l=0}^{j} \dot{\varphi}_\kappa(l/n) \geq \delta \theta/2$. Then,

$$\limsup_{n \to \infty} \bar{\mathrm{E}} \left[\frac{1}{n} \sum_{j=\lceil \delta n \rceil}^{n-1} L \left(\sigma_n \left(\frac{j}{n} \right), \bar{X}_j^n, \dot{\varphi}_\kappa \left(\frac{j}{n} \right) \right); \mathbb{A}^c \cap \left\{ \inf_{j \geq \delta n} \bar{X}_j^n \geq \frac{\delta \theta}{2} \right\} \right]$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{j=\lceil \delta n \rceil}^{n-1} L \left(\sigma_n \left(\frac{j}{n} \right), \frac{1}{n} \sum_{l=0}^{j} \dot{\varphi}_\kappa \left(\frac{l}{n} \right), \dot{\varphi}_\kappa \left(\frac{j}{n} \right) \right). \tag{2.23}$$

Furthermore,

$$\lim_{n \to \infty} \sup_{1 \leq j \leq n} \left| \frac{1}{n} \sum_{l=0}^{j} \dot{\varphi}_\kappa \left(\frac{l}{n} \right) - \int_0^{j/n} \dot{\varphi}_\kappa(s) \, \mathrm{d}s \right| = 0, \tag{2.24}$$

as $\dot{\varphi}_k$ is piecewise constant and bounded. Hence, for all large n, when $\delta n \leq j \leq n$, we have $\delta \theta/4 \leq \int_0^{j/n} \dot{\varphi}_\kappa(s) \, \mathrm{d}s \leq j/n \leq \sigma_n(j/n)(1 - \eta)$. Then, again by the uniform continuity of L on the compact set (2.22), we may analogously bound (2.23) by

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=\lceil \delta n \rceil}^{n-1} L \left(\sigma_n \left(\frac{j}{n} \right), \varphi_\kappa \left(\frac{j}{n} \right), \dot{\varphi}_\kappa \left(\frac{j}{n} \right) \right) = \int_\delta^1 L(\alpha t, \varphi_\kappa(t), \dot{\varphi}_\kappa(t)) \, \mathrm{d}t. \tag{2.25}$$

Step 9. Finally, with respect to the second term in (2.1), by (2.13) and (2.24), in the sup topology, $\lim_{n \to \infty} h(\bar{X}_\cdot^n) = h(\varphi_\kappa(\cdot))$.

We now combine all bounds to conclude the proof of (2.5). By (2.1), and bounds (2.18), (2.20), (2.21), and (2.25), we have

$$\limsup_{n \to \infty} V^n \leq \limsup_{n \to \infty} \bar{\mathrm{E}} \left[\frac{1}{n} \sum_{j=k_1+1}^{n-1} L \left(\sigma_n \left(\frac{j}{n} \right), \bar{X}_j^n, \dot{\varphi}_\kappa \left(\frac{j}{n} \right) \right) + h(\bar{X}_\cdot^n) \right]$$

$$\leq 3\varepsilon + \int_\delta^1 L(\alpha t, \varphi_\kappa(t), \dot{\varphi}_\kappa(t)) \, \mathrm{d}t + h(\varphi_k).$$

Then, by (2.8), (2.9), and (2.10), we obtain (2.5). $\qquad \square$

3. Proofs of Theorems 1.1 and 1.3

As mentioned in the introduction, the differential equation for the pressure is easy to derive heuristically from (1.6); the main technical difficulty is in justifying the convergence of various expressions.

To this end, the key ingredient is the control of the complex zeros of $m_n(z)$, based on (the proof of) [10, Theorem 1].

Proposition 3.1. *Suppose that $m_n(z)$ is defined by (1.5) with initial condition $m_1(\lambda) = \exp\{k_0\lambda\}$ such that $\{s_n\}$ satisfies $s_1 \geq k_0$, $s_n \geq \max\{k_0, 1\}$ for $n \geq 2$. We also assume that there is an $n_0 \geq 1$ such that $s_{n_0} > k_0$ when $k_0 \geq 1$ and that there is an $n_1 \geq 2$ such that $s_{n_1} > 1$ when $k_0 = 0$. Then $m_n(z) \neq 0$ in the strip $|\operatorname{Im}(z)| < \pi$.*

Proof. We note that $m_n(\lambda) = p_n(\mathrm{e}^\lambda)$ for a polynomial $p_n(u)$ with nonnegative coefficients. Then, to prove the result, we recall the proof of [10, Theorem 1] to deduce that all complex zeros of $p_n(u)$ are real, and nonnegative for $n \geq 1$. (Some minor details differ from [10].)

Since $m_n'(\lambda) = up_n'(u)$, (1.5) gives

$$
\begin{aligned}
p_{n+1}(u) &= \frac{u(1-u)}{s_n} p_n'(u) + up_n(u) \\
&= \frac{u}{s_n}(1-u)^{s_n+1} \frac{\mathrm{d}}{\mathrm{d}u}\left((1-u)^{-s_n} p_n(u)\right).
\end{aligned}
\tag{3.1}
$$

Note that $p_1(u) = u^{k_0}$ and that $s_1 \geq k_0$. We first give the proof for the case when $s_1 > k_0 \geq 1$. To reach our conclusion, we prove the following statement by induction.

- For $n \geq 2$, $p_n(u)$ is a polynomial of degree $m_n + k_0$ with a root of multiplicity k_0 at $u = 0$ and $m_n \geq 1$ strictly negative simple roots.

As $s_1 > k_0$, polynomial $p_2(u) = (1/s_1)u^{k_0}((s_1 - k_0)u + k_0)$ satisfies the inductive statement. Suppose that, for some $n \geq 2$, polynomial $p_n(u)$ is of degree $m_n + k_0$, has $m = m_n$ simple negative roots $u_1 < u_2 < \cdots < u_m$, and a root of multiplicity k_0 at $u = 0$. Then, from the first equality in (3.1), $p_{n+1}(u)$ also has a root of multiplicity k_0 at $u = 0$. Clearly, $\{u_1, u_2, \ldots, u_m, 0\}$ are m distinct roots of the expression under the derivative on the right-hand side of (3.1). Since $\{u_j\}$ are simple roots, the function must cross the real line, so by Rolle's theorem, $p_{n+1}(u)$ has m distinct roots interlaced between the roots of $p_n(u)$. This shows that $p_{n+1}(u)$ has $m + k_0$ real roots in the interval $(u_1, 0]$. This ends the proof if the degree of p_{n+1} is $m + k_0$, which occurs when $s_n = m$.

If $s_n \neq m$ then the degree of p_{n+1} is $m + 1 + k_0$, and to end the proof, we want to show that the last $(m + k_0 + 1)$th root of $p_{n+1}(u)$ is located to the left of u_1, so that all negative roots of $p_{n+1}(u)$ must be simple. To see this, we again follow [10]. The first equation in (3.1) shows that $p_{n+1}(u_1)$ and $p_n'(u_1)$ have opposite signs, and that $p_n'(u_1) \neq 0$ as u_1 is a simple root. So $p_{n+1}(u_1) \neq 0$. Since polynomials $p_{n+1}(u)$ and $p_n(u)$ have positive leading coefficients and their degrees differ by 1, their signs are opposite as $u \to -\infty$. Since u_1 is the smallest root, $p_n(u)$ has constant sign for $u < u_1$, which matches the sign of $p_{n+1}(u_1)$. Thus, $p_{n+1}(u)$ must eventually cross the real line to the left of u_1. This shows that $p_{n+1}(u)$ has $m + 1$ simple strictly negative roots, and a root of multiplicity k_0 at $u = 0$, ending the induction proof.

Next, suppose that $s_1 = s_2 = s_{n_0-1} = k_0 \geq 1$, but $s_{n_0} > k_0$ for some $n_0 \geq 2$. Then $p_1(u) = p_2(u) = \cdots = p_{n_0}(u) = u^{k_0}$ and the inductive proof goes through with minor modifications, starting with $p_{n_0+1}(u) = (1/s_{n_0})u^{k_0}((s_{n_0} - k_0)u + k_0)$ that replaces $p_2(u)$ in the previous argument.

Finally, if $k_0 = 0$ then $p_1(u) = 1$ and $p_2(u) = u$. First choose $n_1 \geq 2$ such that $s_{n_1} > 1$ but $s_n = 1$ for $2 \leq n < n_1$. Since in this case the value of s_1 is irrelevant, we obtain $p_2(u) = \cdots = p_{n_1}(u) = u$. The induction proof proceeds with minor modifications, starting with $p_{n_1+1}(u) = (1/s_{n_1})u((s_{n_1} - 1)u + 1)$.

Proposition 3.2. *Under the assumptions of Proposition 3.1, suppose that*

$$\frac{1}{n}\log m_n(\lambda) \to \Lambda(\lambda) \quad as\ n \to \infty. \tag{3.2}$$

Then $\Lambda(\lambda)$ is analytic, satisfies (1.9), and is given by (1.13).

The main step in the proof is the following.

Lemma 3.1. *Let $\Lambda(\lambda)$ be as in Proposition 3.2. Then Λ is analytic and*

$$\lim_{n\to\infty} \frac{m_n'(\lambda)}{nm_n(\lambda)} = \Lambda'(\lambda).$$

Proof. Recall that if $f(\zeta)$ is holomorphic in $|\zeta| \le R$ then, for $0 < r < R$,

$$\max_{|\zeta|\le r} |f(\zeta)| \le \frac{R+r}{R-r}|f(0)| + \frac{2r}{R-r} \max_{|\zeta|\le R} \operatorname{Re} f(\zeta) \tag{3.3}$$

(cf. [14, Theorem 6.31(ii)]).

By Proposition 3.1, for fixed $n \ge 1$, the function $m_n(z)$ is holomorphic and nonzero in the strip $|\operatorname{Im} z| < \pi$. Since $m_n(z)$ is a polynomial in e^z with nonnegative coefficients, $m_n(t) > 0$ for all $t \in \mathbb{R}$, $m_n(t)$ is increasing on $\mathbb{R}$, and $|m_n(z)| \le m_n(|z|)$.

The function $m_n(z)$ has a holomorphic logarithm $L_n(z)$ on the strip $|\operatorname{Im} z| < \pi$. Because $m_n(t) > 0$ for $t \in \mathbb{R}$, we may assume that $L_n(t) = \log(m_n(t))$ for all $t \in \mathbb{R}$. For each $t \in \mathbb{R}$, we can apply (3.3) to $f(\zeta) = n^{-1}L_n(t+\zeta)$, $|\zeta| < \pi$, with $R = 2\pi/3$ and $r = \pi/3$. This gives

$$\max_{|z-t|\le \pi/3} n^{-1}|L_n(z)| \le 3|n^{-1}L_n(t)| + 2n^{-1} \max_{|z-t|<\pi/3} |\operatorname{Re} L_n(\zeta)|$$

$$= 3|n^{-1}\log(m_n(t))| + 2n^{-1} \max_{|z-t|<\pi/3} \log|m_n(z)|$$

$$\le 3|n^{-1}\log(m_n(t))| + 2\log\left(m_n\left(|t| + \frac{2\pi}{3}\right)\right).$$

Since $e^{-n|t|} \le m_n(t) \le e^{n|t|}$, $\{n^{-1}L_n(z)\colon n \ge 1\}$ is a normal family (i.e. a uniformly bounded family of holomorphic functions) in the disk $|z - t| \le \pi/3$.

We now note that $\{n^{-1}L_n(t)\}$ converges for all real t by (3.2). A version of Vitalli's theorem (see [22, p. 9]) implies that $n^{-1}L_n(z) \to \Lambda(z)$ in the strip $|\operatorname{Im} z| \le \pi/3$, the convergence is uniform in each disc $|z - t| \le \pi/3$, the limit $\Lambda(z)$ is an analytic function of the argument z in that strip, and all derivatives of $n^{-1}L_n$ converge to the corresponding derivatives of Λ. In particular, the sequence $n^{-1}m_n'(\lambda)/m_n(\lambda) = n^{-1}L_n'(\lambda)$ converges to $\Lambda'(\lambda)$ for all real λ.

Proof of Proposition 3.2. By Lemma 3.1, the right-hand side of (1.6) converges, so the left-hand side must converge too: $m_{n+1}(\lambda)/m_n(\lambda) \to e^{L(\lambda)}$ for some L uniformly in a neighborhood of λ. Since the limit of ratios implies the same limit for nth roots, we obtain $(1/n)\log m_n(\lambda) \to L(\lambda)$, which identifies $L(\lambda) = \Lambda(\lambda)$ as the pressure. From Lemma 3.1, the derivative $m_n'(\lambda)/(nm_n(\lambda)) \to \Lambda'(\lambda)$, so passing to the limit in (1.6) we obtain the differential equation for the pressure, (1.9).

It is straightforward to verify that the expression on the right-hand side of (1.13) satisfies (1.9). Therefore, to end the proof, it remains to show that the initial value problem $\Lambda(0) = 0$ for differential equation (1.9) has a unique solution. We proceed by contradiction. Suppose that

Λ_1 and Λ_2 are two different solutions on $(0, \infty)$ with initial condition $\Lambda(0) = 0$. If $\Lambda_1(t) = \Lambda_2(t)$ for some $t > 0$ then they coincide for all $t > 0$. Therefore, one of the solutions, say Λ_1, is above the other and we must have $\Lambda_1(t) > \Lambda_2(t)$ for all $t > 0$. By the mean value theorem, there is a $t_0 > 0$ such that $\Lambda'_1(t_0) - \Lambda'_2(t_0) > 0$. But the equation gives $\Lambda'_1(t) - \Lambda'_2(t) = \alpha(\exp\{\Lambda_1(t)\} - \exp\{\Lambda_2(t)\})/(1 - e^t) < 0$ for all $t > 0$. This contradiction shows that there is only one solution on $(0, \infty)$. Similarly, suppose that Λ_1 and Λ_2 are two different solutions on $(-\infty, 0)$ with initial condition $\Lambda(0) = 0$. If $\Lambda_1(t) = \Lambda_2(t)$ for some $t < 0$ then they coincide for all $t < 0$. Therefore, one of the solutions, say Λ_1, is above the other and we must have $\Lambda_1(t) > \Lambda_2(t)$ for all $t < 0$. By the mean value theorem, there is a $t_0 < 0$ such that $\Lambda'_1(t_0) - \Lambda'_2(t_0) = (\Lambda_1(t) - \Lambda_2(t))/t < 0$. But the equation gives $\Lambda'_1(t) - \Lambda'_2(t) = \alpha(\exp\{\Lambda_1(t)\} - \exp\{\Lambda_2(t)\})/(1 - e^t) > 0$ for all $t < 0$, a contradiction. Thus, the solution is unique on $(-\infty, 0)$ too.

Our final step is to prove that limit (3.2) exists under (1.3) or (1.4).

Proposition 3.3. *If either (1.3) or (1.4) holds, then limit (3.2) exists, and is given by the smooth function*

$$\Lambda(\lambda) = \begin{cases} \log\left(\dfrac{\sqrt{e^\lambda - 1}}{\arctan(\sqrt{e^\lambda - 1})}\right) & \text{if } \lambda > 0, \\[3mm] \log\left(\dfrac{\sqrt{1 - e^\lambda}}{\text{arctanh}(\sqrt{1 - e^\lambda})}\right) & \text{if } \lambda < 0, \end{cases} \quad \text{for } \alpha = \tfrac{1}{2}, \tag{3.4}$$

$$\Lambda(\lambda) = \log\frac{e^\lambda - 1}{\lambda} \quad \text{for } \alpha = 1, \tag{3.5}$$

when $\lambda \neq 0$, and $\Lambda(0) = 0$.

Proof. We follow analytic arguments adapted from [25]. Since $0 \leq k_0 \leq Z_n \leq n + k_0 - 1$, we obtain $0 < m_n(\lambda) \leq \exp\{(n + k_0)\lambda^+\}$ with $\lambda^+ = \max\{\lambda, 0\}$. Therefore, for all complex z with $|z| < e^{-\lambda^+}$, $G(z, \lambda)$ is well defined and satisfies (1.10) with the initial condition $G(0, \lambda) = \exp\{k_0\lambda\}$. The coefficients of this PDE do not vanish in the regions

$$\mathcal{D}_+ = \{(z, \lambda) \colon \lambda > 0, \ |z| < e^{-\lambda}\}, \qquad \mathcal{D}_- = \{(z, \lambda) \colon \lambda < 0, \ |z| < 1\}.$$

For $\lambda \neq 0$, the PDE can be solved by the method of characteristics. Clearly, $m_n(\lambda)$ are the coefficients in the Taylor expansion at $z = 0$ of the solution, and $e^{-\Lambda(\lambda)}$ is the radius of convergence of the series that can be determined by singularity analysis.

For $\alpha = 1$ and $k_0 = 0, 1$, using the initial condition $G(0, \lambda) = \exp\{k_0\lambda\}$, we obtain

$$G(\lambda, z) = \exp\{k_0(z(e^\lambda - 1) - \lambda)\}\frac{e^\lambda - 1}{1 - \exp\{z(e^\lambda - 1) - \lambda\}}.$$

Hence, the singularity of G as a function of z nearest to the origin is a simple pole at

$$z_0 = \frac{\lambda}{e^\lambda - 1}.$$

By Darboux's asymptotic method [38, Chapter 8],

$$\frac{1}{n}\log E[\exp\{\lambda Z_n\}] \to \log\frac{e^\lambda - 1}{\lambda}$$

and (3.5) follows.

Next, consider $\alpha = \frac{1}{2}$. In this case, the solutions of the PDE differ depending on the region $\mathcal{D}_{\pm}$, but are explicit so there are no difficulties in constructing their analytic extensions. Using the initial condition $k_0 = 0$, we have $G(0, \lambda) = 1$, and the solution of (1.10) is

$$
G(\lambda, z) = \begin{cases} \dfrac{\sqrt{e^\lambda - 1}}{\tan(\arctan \sqrt{e^\lambda - 1} - z\sqrt{e^\lambda - 1})}, & \lambda > 0, \\[3mm] \dfrac{\sqrt{1 - e^\lambda}}{\tanh(\operatorname{arctanh}\sqrt{1 - e^\lambda} - z\sqrt{1 - e^\lambda})}, & \lambda < 0. \end{cases}
$$

Hence, the singularity of G as a function of z nearest to the origin is a simple pole at

$$
z_0 = \begin{cases} \dfrac{\arctan \sqrt{e^\lambda - 1}}{\sqrt{e^\lambda - 1}}, & \lambda > 0, \\[3mm] \dfrac{\operatorname{arctanh}\sqrt{1 - e^\lambda}}{\sqrt{1 - e^\lambda}}, & \lambda < 0. \end{cases}
$$

Once again, by Darboux's asymptotic method [38, Chapter 8],

$$
\frac{1}{n} \log \mathrm{E}[\exp\{\lambda Z_n\}] \to \log \frac{1}{z_0}
$$

and (3.4) follows.

Proof of Theorem 1.1. We first observe that s_n, given by (1.2), (1.3), or (1.4), satisfies the assumptions of Proposition 3.2, that is, those inherited from Proposition 3.1.

Next, we verify that assumption (3.2) of Proposition 3.2 is also satisfied. To do so, we first consider the case when condition (1.2) holds. Since

$$
\frac{Z_n}{n} = \frac{n + k_0 - 1}{n} X_{n+k_0-1}(1),
$$

by the contraction principle, Theorem 1.2 implies Theorem 1.1, except for the form of the rate function I. As $x \mapsto \lambda x$ is a bounded continuous function on $[0, 1]$, by Varadhan's integral lemma (see [17, Theorem 4.3.1] or [20, Theorem 1.3.4]), this implies that limit (1.12) exists. Thus, (3.2) holds in this case.

Finally, when either (1.3) or (1.4) holds, limit (3.2) exists by Proposition 3.3.

The LDP with rate function I, noting smoothness and convexity of Λ, is now obtained by the Gärtner–Ellis theorem [17, Theorem 2.3.6] from Proposition 3.2.

The strict convexity of I on its domain follows from the strict convexity of Λ on $(-\infty, \infty)$. The latter is most easily seen from (1.9), which has no solutions that are linear on an interval.

Proof of Theorem 1.3. For the CLT, we recall [13, Proposition 2] in our context. When $\sup_n m_n(\epsilon)^{1/n} < \infty$ for some $\epsilon > 0$, $0 \notin \text{closure}(\bigcup_{n \geq 1} Z_{e_n})$, where Z_{e_n} is the zero set of $m_n(z) = \mathrm{E}[\exp\{z Z_n\}]$, and Z_n/n satisfies an LDP, then $(Z_n - \mathrm{E}[Z_n])/\sqrt{n}$ converges in distribution to $N(0, \sigma^2)$, where $\sigma^2 = \Lambda''(0)$.

To verify the assumptions of [13, Proposition 2], note that as part of the proof of Theorem 1.1 we have already verified convergence (3.2) for all $\lambda \in \mathbb{R}$, and that, by Proposition 3.1, $m_n(z)$ has no zeros in the strip $|\operatorname{Im}(z)| < \pi$ for $n \geq 1$. Finally, we may use (1.9) to verify that $\Lambda''(0) = \alpha^2/[(1+\alpha)^2(2+\alpha)]$ to finish the proof.

Also, $\Lambda'(0) = \alpha/(1+\alpha)$, and the LLN follows from the strict convexity of I near $\alpha/(1+\alpha)$ in Theorem 1.1, given that Λ is analytic and $\Lambda''(0) > 0$.

4. Singularity analysis for $s_n = 2n$

Formula (1.13) was in fact obtained by us from heuristic singularity analysis for $s_n = \alpha n$; once (1.13) is available, it is easy to verify that it gives the solution of ODE (1.9). Since the singularity technique [25] is virtually unknown to probabilists, it is therefore of interest to provide a nontrivial example.

The following result is a version of Proposition 3.3 and can also be derived by explicit evaluation of the integral in (1.13).

Proposition 4.1. *If $s_n = 2n$ then limit (1.12) (or (3.2)) exists, and is given by the smooth function*

$$\Lambda(\lambda) = \log\left(\frac{(e^\lambda - 1)^2}{2(e^\lambda - 1 - \lambda)}\right) \tag{4.1}$$

when $\lambda \neq 0$, and $\Lambda(0) = 0$.

For $\alpha = 2$, the method of characteristics gives the following solution of (1.10).

Lemma 4.1. *Suppose that φ is a function of one complex variable, analytic in a domain $\mathcal{D}$ containing $(-\infty, -2)$. Then*

$$G(z, \lambda) := (e^\lambda - 1)^2 \varphi(z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda)$$

satisfies PDE (1.10) for all $\lambda \neq 0$ and small enough $|z|$ with $\alpha = 2$. Furthermore, the initial condition is fulfilled at $\lambda \neq 0$ if

$$\varphi(2\lambda - 2e^\lambda) = \frac{\exp\{k_0\lambda\}}{(1 - e^\lambda)^2}. \tag{4.2}$$

Proof. The verification of the initial condition is trivial, and the verification of the PDE is a straightforward calculation. Denoting, for conciseness, $\psi(z, \lambda) = \varphi(z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda)$ and $\psi^{(1)}(z, \lambda) = \varphi'(z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda)$, we verify, for $\lambda \neq 0$ and z such that $z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda \in \mathcal{D}$, that

$$\frac{\partial G(z, \lambda)}{\partial \lambda} = 2(e^\lambda - 1)e^\lambda \psi(z, \lambda) + 2(e^\lambda - 1)^3(ze^\lambda - 1)\psi^{(1)}(z, \lambda)$$

and

$$\frac{\partial G(z, \lambda)}{\partial z} = (e^\lambda - 1)^4 \psi^{(1)}(z, \lambda).$$

Equation (1.10) now follows by a calculation.

Our next goal is to show that we can find a solution of (1.10) which can be analytically extended in variable z to a large enough domain. To this end, we need to analyze (4.2) for complex λ. The basic plan consists of noting that the function $f(x) = 2(x - e^x)$ is analytic, strictly decreasing for $x > 0$, strictly increasing for $x < 0$, and $f(0) = -2$. The derivative $f'(x)$ vanishes only at $x = 0$, so $f \mid_{[0,\infty)}$ and $f \mid_{(-\infty,0]}$ have continuous inverses $h_+ : (-\infty, -2] \to [0, \infty)$ and $h_- : (-\infty, -2] \to (-\infty, 0]$, and both are analytic on $(-\infty, -2)$.

Clearly, if we define

$$\varphi_\pm(\lambda) = \frac{\exp\{k_0 h_\pm(\lambda)\}}{(\exp\{h_\pm(\lambda)\} - 1)^2},$$

then φ_+ satisfies (4.2) for $\lambda > 0$, and φ_- satisfies (4.2) for $\lambda < 0$. The goal is therefore to find the appropriate analytic extensions of the functions $h_\pm$.
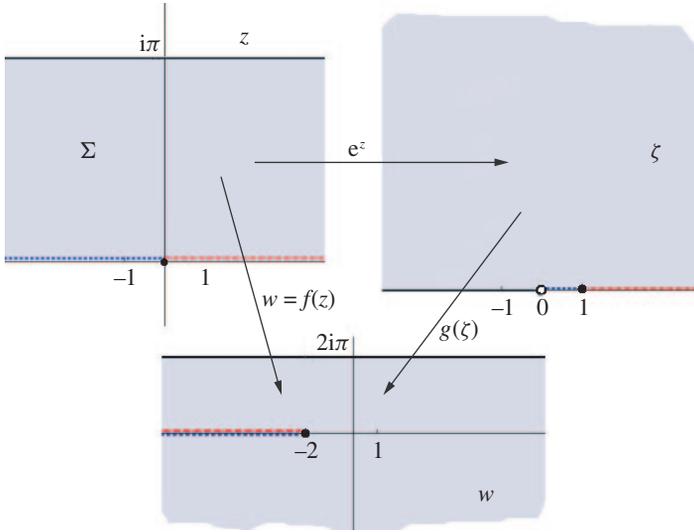
FIGURE 2: Composition of maps $f(z) = g(e^z)$ for the proof of Lemma 4.3.

To accomplish this goal, we need to analyze $f(z) = 2(z - e^z)$. The closely related function $z + e^z$ appears in [33, p. 116], but proofs are not included there; we give details for $f(z) = 2(z - e^z)$ for completeness.

The proof relies on the following univalence criterion.

Lemma 4.2. (Wolff–Warschawski–Nishiro.) *If g is holomorphic in a convex region Ω and $g'(\Omega) \subset H$, a half-plane with $0 \in \partial H$, then g is one-to-one on Ω.*

Proof. This is [22, Theorem 2.16, p. 47] applied to the function $e^{i\theta} g(z)$ with a real constant θ chosen appropriately to rotate H.

Lemma 4.3. *Let f be a one-to-one mapping of the half-closed strip $\Sigma = \{z: 0 < \text{Im}(z) \leq \pi\}$ onto a slit closed half-plane: $\{w: \text{Im}(w) \leq \pi\} \setminus (-\infty, -2]$. The boundary correspondence is as follows: f maps $\mathbb{R} + \pi i$ injectively onto $\mathbb{R} + 4\pi i$, and f is one-to-one on both $[0, +\infty)$ and $(-\infty, 0]$, and maps each onto $(-\infty, -2]$ (cf. Figure 2).*

Proof. We write $f(z) = g(\zeta)$, where $\zeta = e^z$, $g(\zeta) = 2(\log \zeta - \zeta)$ and log denotes the principal branch of the logarithm. The function $\zeta = e^z$ is a one-to-one mapping on the strip Σ. The image of the interior of Σ is the upper half-plane $\mathbb{H} = \{\zeta: \text{Im}(\zeta) > 0\}$. Furthermore, the upper edge $\mathbb{R} + \pi i$ of Σ is mapped onto $(-\infty, 0)$ and the bottom edge $\mathbb{R}$ is carried onto $(0, +\infty)$.

The derivative of $g(\zeta)$ is $g'(\zeta) = 2\bar{\zeta}/|\zeta|^2 - 2$. In particular,

$$\text{Im}(g'(\zeta)) = -\frac{2\,\text{Im}(\zeta)}{|\zeta|^2} < 0 \quad \text{for } \zeta \in \mathbb{H}.$$

By Lemma 4.2 with $\Omega = \mathbb{H}$, g is one-to-one on the half-plane $\mathbb{H}$. Under g, the image of $\mathbb{H}$ is $\{w: \text{Im}(w) \leq 2\pi\} \setminus (-\infty, -2]$, and on the boundary we have $g(\mathbb{R}^-) = \mathbb{R} + 2\pi i$ and $g(\mathbb{R}^+)$ is the slit $(-\infty, -2]$ twice covered with $g(1) = -2$. Perhaps this latter statement is easier seen directly from f; the slit is covered twice: $f((-\infty, 0]) = f([0, +\infty)) = (-\infty, -2]$.
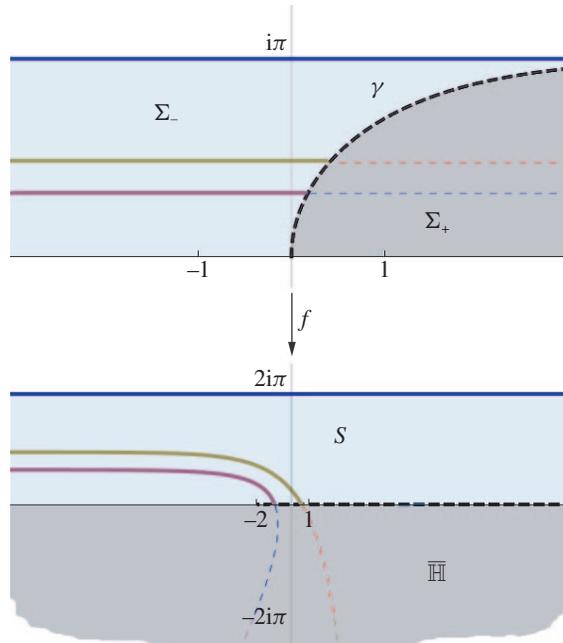
FIGURE 3: Conformal maps $f \mid \Sigma_-$ and $f \mid \Sigma_+$.

We investigate the conformal mapping $f \mid \Sigma$ in more detail. The preimage of $[-2, \infty)$ under $f \mid \Sigma$ is the curve γ given by $\operatorname{Re} z = \log(\operatorname{Im} z / \sin \operatorname{Im} z)$, $0 \leq \operatorname{Im} z < \pi$. This curve begins at the origin and becomes asymptotic to $\mathbb{R} + \pi \mathrm{i}$ in the positive direction. By removing the curve γ, Σ is cut into two parts, Σ_- and Σ_+. The region Σ_- is bounded by $\mathbb{R}^-$, γ, and $\mathbb{R} + \pi \mathrm{i}$, with the latter line part of Σ_-. The region Σ_+ is bounded by γ and $\mathbb{R}^+$. Then $f \mid \Sigma_-$ is a conformal mapping of Σ_- onto the half-closed strip $S = \{w : 0 < \operatorname{Im} w \leq 2\pi\}$ with $f(\mathbb{R} + \pi \mathrm{i}) = \mathbb{R} + 2\pi \mathrm{i}$, $f(\mathbb{R}^-) = (-\infty, -2)$, and $f(\gamma) = [-2, +\infty)$. Similarly, $f \mid \Sigma_+$ is a conformal mapping of Σ_+ onto the lower half-plane $\overline{\mathbb{H}} = \{\bar{z} : z \in \mathbb{H}\}$, where $\bar{z}$ denotes the complex conjugate of z, with $f(\mathbb{R}^+) = (-\infty, -2)$ and $f(\gamma) = [-2, +\infty)$. (See Figure 3.)

Both maps $f \mid \Sigma_-$ and $f \mid \Sigma_+$ extend to conformal maps of larger regions. Because $f(\bar{z}) = \overline{f(z)}$, f maps $\overline{\Sigma_+}$ conformally onto $\mathbb{H}$. As $f(\mathbb{R}^+) = (-\infty, -2)$, f is a conformal map of $\Omega_+ := \Sigma_+ \cup \mathbb{R}^+ \cup \overline{\Sigma_+}$ onto the slit plane $\overline{\mathbb{H}} \cup (-\infty, -2) \cup \mathbb{H} = \mathbb{C} \setminus [-2, +\infty)$. Let $h_+ : \mathbb{C} \setminus [-2, +\infty) \to \Omega_+$ be the inverse function for $f \mid \Omega_+$. The conformal extension of $f \mid \Sigma^-$ is more involved to describe. The fact that $f(\bar{z}) = \overline{f(z)}$ implies that f maps $\overline{\Sigma_-}$ conformally onto $\bar{S}$. Since $f(\mathbb{R}^-) = (-\infty, -2)$, f is a conformal map of $\Delta_0 = \Sigma_- \cup \mathbb{R}^- \cup \overline{\Sigma_-}$ onto the slit closed strip $S_0 = S \cup (-\infty, -2) \cup \bar{S} = \{w : |\operatorname{Im} w| \leq 2\pi\} \setminus [-2, +\infty)$ with the upper and lower edges of Δ_0 respectively corresponding to the upper and lower edges of S_0. It is straightforward to verify that $f(z + 2\pi \mathrm{i}) = f(z) + 4\pi \mathrm{i}$ for each $n \in \mathbb{Z}$. This functional relationship implies that f is a conformal map of $\Delta_n = \Delta_0 + 2\pi n \mathrm{i}$ onto $S_n = S_0 + 4\pi n \mathrm{i}$ for each $n \in \mathbb{Z}$. Hence, f is a conformal map of $\Omega_- := \bigcup_{n \in \mathbb{Z}} \Delta_n$ onto the infinitely slit plane $S_\infty := \bigcup_{n \in \mathbb{Z}} S_n = \mathbb{C} \setminus \bigcup_{n \in \mathbb{Z}}([-2, +\infty) + 4\pi n \mathrm{i})$. Let $h_- : S_\infty \to \Omega_-$ be the inverse function of $f \mid \Omega_-$. Because $\mathbb{C} \setminus [-2, +\infty) \supset S_\infty$, we may regard h_+ as defined on S_∞, so $h_\pm$ have a common domain. (See Figure 4.)
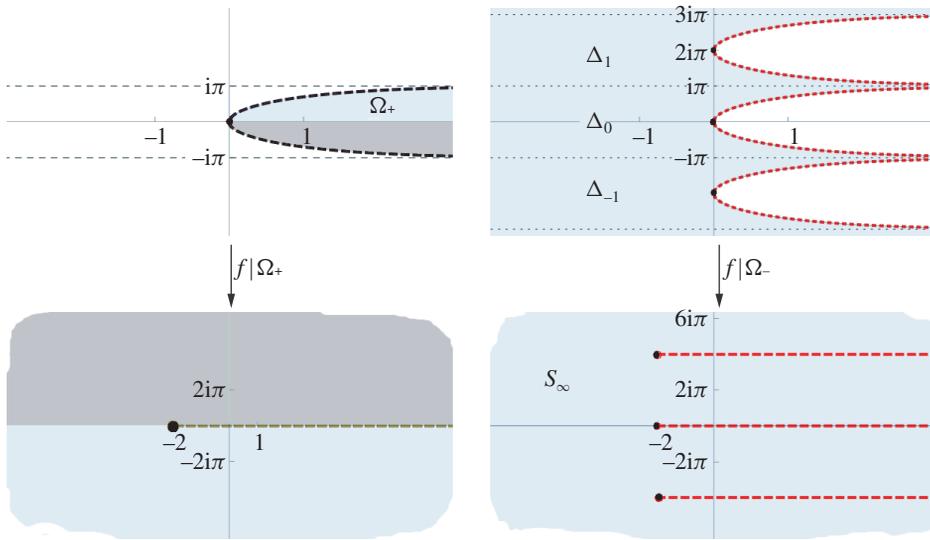
FIGURE 4: Conformal maps $f \mid \Omega_+$ and $f \mid \Omega_-$ with $\Omega_+ := \Sigma_+ \cup \mathbb{R}^+ \cup \overline{\Sigma}_+$ and $\Omega_- = \bigcup_{n=-\infty}^{\infty} \Delta_n$.

The two conformal maps $h_\pm$ just constructed provide analytic extensions to S_∞ of the real-valued functions $h_\pm$. This allows us to define a pair of functions

$$\varphi_\pm(z) := \frac{\exp\{k_0 h_\pm(z)\}}{(\exp\{h_\pm(z)\} - 1)^2},$$

which are analytic in S_∞.

Lemma 4.4. *For each point $u \in (-\infty, -2)$, the power series expansion of $\varphi_\pm$ has radius of convergence $2 - u$, $\varphi_\pm$ is analytic in the slit disk $D(u, r) \setminus [-2, +\infty)$, where $r = r(u) = \sqrt{(u + 2)^2 + 16\pi^2}$. Furthermore, $\varphi_\pm(w) \approx -1/(w + 2)$ as $w \to -2$ in S_∞.*

Proof. For each point $u \in (-\infty, -2)$, the power series expansion of $h_\pm$ has radius of convergence $|u + 2|$, the distance from u to -2 because -2 is the closest singularity of $h_\pm$. Also, $h_\pm$ is analytic in the slit disk $D(u, r) \setminus [-2, +\infty)$, where $r = r(u) = \sqrt{(u + 2)^2 + 16\pi^2}$ is the distance from u to $-2 \pm 4\pi \mathrm{i}$.

Note that singularities of φ_- that arise from $h_-(w) = 2n\pi \mathrm{i}$ are located at the slits taken out of S_∞ and that $h_+(w) \in \Omega_+$ cannot take values in $2n\pi \mathrm{i}$. Thus, $\varphi_\pm$ is also analytic in the slit disk $D(u, r) \setminus [-2, +\infty)$.

Substituting $w = 2(z - \mathrm{e}^z)$, we see that

$$
\begin{aligned}
\lim_{\substack{w \to -2 \\ w \in S_\infty}} (w + 2)\varphi_\pm(w) &= \lim_{\substack{w \to -2 \\ w \in S_\infty}} \exp\{k_0 h_\pm(w)\} \frac{w + 2}{(\exp\{h_\pm(w)\} - 1)^2} \\
&= 2 \lim_{\substack{z \to 0 \\ z \in \Omega_\pm}} \exp\{k_0 z\} \frac{1 + z - \mathrm{e}^z}{(\mathrm{e}^z - 1)^2} \\
&= -1.
\end{aligned}
$$

Proof of Proposition 4.1. We now prove (1.12) and identify the limit. Using functions $\varphi_\pm$ constructed above, define

$$\widetilde{G}(z, \lambda) := \begin{cases} (e^\lambda - 1)^2 \varphi_+(z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda) & \text{if } \lambda > 0, \\ (e^\lambda - 1)^2 \varphi_-(z(e^\lambda - 1)^2 + 2\lambda - 2e^\lambda) & \text{if } \lambda < 0. \end{cases}$$

From Lemma 4.1 we see that the function $\widetilde{G}(z, \lambda)$ satisfies (1.10) for all $(z, \lambda) \in \mathcal{D}_+ \cup \mathcal{D}_-$. By uniqueness of the PDE solution in each of the two regions, $G(z, \lambda) = \widetilde{G}(z, \lambda)$ for all $(z, \lambda) \in \mathcal{D}_+ \cup \mathcal{D}_-$. In particular, for each fixed $\lambda \neq 0$, $\lim \sup_{n \to \infty} (m_n(\lambda))^{1/n}$ is the reciprocal of the radius of convergence of the series expansion of $\widetilde{G}(z, \lambda)$ at $z = 0$. The latter is $(1 - e^\lambda)^{-2}$ times the radius of convergence for $\varphi_{\mathrm{sgn}(\lambda)}(w)$ at $u = 2\lambda - 2e^\lambda \in (-\infty, -2)$, which by Lemma 4.4 is $2(e^\lambda - \lambda - 1)$. Furthermore, the lemma implies that there is an $\eta = \eta(\lambda) > 0$ such that $\varphi_{\mathrm{sgn}(\lambda)}(w)$ is analytic on the slit disk $\{w : |w - u| < 2(e^\lambda - \lambda - 1)(1 + \eta)\} \setminus [-2, \infty)$. Since after appropriate translation and rescaling this slit disk is larger than the indented disk $\Delta(\pi/4, \eta)$ introduced in [24, Equation (2.5)], and the second part of Lemma 4.4 gives $\varphi_\pm(w) \approx -1/(w+2)$ as $w \to -2$, we can apply [24, Corollary 2] to obtain (1.12).

Finally, as $m_n(0) = 1$, the convergence at $\lambda = 0$ holds trivially. $\qquad \blacksquare$

5. Concluding remarks

1. *Variational methods.* By Varadhan's integral lemma, Theorem 1.2 implies that limit (1.12) exists, and equals

$$\Lambda(\lambda) = \sup_\varphi \{\lambda\varphi(1) - I(\varphi)\}$$

$$= \sup_{\{\varphi : \varphi(0)=0\}} \left\{ \int_0^1 \left[\lambda\dot{\varphi}(t) - \dot{\varphi}(t) \log \frac{\alpha t \dot{\varphi}(t)}{\alpha t - \varphi(t)} \right. \right.$$

$$\left. \left. - (1 - \dot{\varphi}(t)) \log \frac{\alpha t (1 - \dot{\varphi}(t))}{\varphi(t)} \right] dt \right\}. \tag{5.1}$$

Direct derivation of (1.13), or even (4.1) from (5.1) seems quite challenging (cf. [44]).

2. *Cases $\alpha = \frac{1}{2}, 1$.* Although we prove a LDP for Z_n/n when $s_n = n$ or $s_2 = n/2$ (Theorem 1.1), the proof of Theorem 1.3 for a path LDP with respect to $Z_{\lfloor nt \rfloor}/n$, especially the lower bound argument, does not cover these cases. The difficulty is in controlling the boundary behavior as estimate (2.11) is not available. In particular, it would be interesting to investigate if Theorem 1.1 holds true when assumption (1.3) is relaxed to $s_n/n \to 1$.

3. *Higher-order statistics.* With respect to random graph models, one might ask about LDPs for the vector $\boldsymbol{T}_n^k/n = \langle T_n(1), \ldots, T_n(k) \rangle / n$, where the jth component $T_n(j)$ counts the number of vertices with degree $j \leq k$ for $k \geq 2$. In principle, our method to analyze the leaves can be used to study $\boldsymbol{T}_n^k/n$. Indeed, the Dupuis–Ellis-type arguments given here for a path LDP for the leaves $T_{\lfloor nt \rfloor}(1)/n$ (Proposition 1.1) would seem to extend to the vector-valued paths $\boldsymbol{T}_{\lfloor nt \rfloor}^k/n$.

However, to calculate the pressure

$$\boldsymbol{\Lambda}_k(\lambda_1, \ldots, \lambda_k) = \lim_{n \to \infty} \frac{1}{n} \log \mathrm{E}\left[\exp\left\{ \sum_{i=1}^k \lambda_i T_n(i) \right\} \right],$$

as in Theorem 1.2 with respect to the leaves, the differential equation which now arises for $\boldsymbol{\Lambda}_k$, in place of the ODE for Λ (1.9), is a quasilinear PDE with $k \geq 2$ independent variables. These

PDEs, although in principle implicitly solved by the method of characteristics, unfortunately do not seem to admit explicit solutions, at least to the extent found here with respect to $T_n(1)/n$, a reason why we have focused on detailed investigations of the leaves. It would be of interest to better study these higher-order questions.

Acknowledgements

References

[1] ALBERT, R. AND BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74,** 47–97.

[2] ALDOUS, D. J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statist. Sci.* **16,** 23–34.

[3] ATHREYA, K. B., GHOSH, A. P. AND SETHURAMAN, S. (2008). Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci.* **118,** 473–494.

[4] BAKHTIN, Y. AND HEITSCH, C. (2008). Large deviations for random trees. *J. Statist. Phys.* **132,** 551–560.

[5] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[6] BENASSI, A. (1996). Arbres et grandes déviations. In *Trees* (Versailles, 1995; Progress Prob. **40**), Birkhäuser, Basel, pp. 135–140.

[7] BLUM, M. G. B. AND FRANÇOIS, O. (2005). On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. *Math. Biosci.* **195,** 141–153.

[8] BLUM, M. G. B., FRANÇOIS, O. AND JANSON, S. (2006). The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Prob.* **16,** 2195–2214.

[9] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. AND TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18,** 279–290.

[10] BONA, M. (2008/09). Real zeros and normal distribution for statistics on Stirling permutations defined by Gessel and Stanley. *SIAM J. Discrete Math.* **23,** 401–406.

[11] BONATO, A. (2008). *A Course on the Web Graph* (Graduate Stud. Math. **89**). American Mathematical Society, Providence, RI.

[12] BROUTIN, N. AND DEVROYE, L. (2006). Large deviations for the weighted height of an extended class of trees. *Algorithmica* **46,** 271–297.

[13] BRYC, W. (1993). A remark on the connection between the large deviation principle and the central limit theorem. *Statist. Prob. Lett.* **18,** 253–256.

[14] BURCKEL, R. B. (1979). *An Introduction to Classical Complex Analysis*. Vol. 1 (Pure Appl. Math. **82**) Academic Press, New York.

[15] CHUNG, F. AND LU, L. (2006). *Complex Graphs and Networks* (CBMS Regional Conf. Ser. Math. **107**). American Mathematical Society, Providence, RI.

[16] COOPER, C. AND FRIEZE, A. (2003). A general model of web graphs. *Random Structures Algorithms* **22,** 311–335.

[17] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. (Appl. Math. **38**). Springer, New York.

[18] DEMBO, A., MÖRTERS, P. AND SHEFFIELD, S. (2005). Large deviations of Markov chains indexed by random trees. *Ann. Inst. H. Poincaré Prob. Statist.* **41,** 971–996.

[19] DEREICH, S. AND MÖRTERS, P. (2009). Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Prob.* **14,** 1222–1267.

[20] DUPUIS, P. AND ELLIS, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.

[21] DUPUIS, P., NUZMAN, C. AND WHITING, P. (2004). Large deviation asymptotics for occupancy problems. *Ann. Prob.* **32,** 2765–2818.

[22] DUREN, P. L. (1983). *Univalent Functions*. Springer, New York.

[23] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge University Press.

[24] FLAJOLET, P. AND ODLYZKO, A. (1990). Singularity analysis of generating functions. *SIAM J. Discrete Math.* **3,** 216–240.

[25] FLAJOLET, P., GABARRÓ, J. AND PEKARI, H. (2005). Analytic urns. *Ann. Prob.* **33,** 1200–1233.

[26] FRANÇOIS, O. AND MIOLAND, C. (2007). Gaussian approximations for phylogenetic branch length statistics under stochastic models of biodiversity. *Math. Biosci.* **209,** 108–123.

[27] GESSEL, I. AND STANLEY, R. P. (1978). Stirling polynomials. *J. Combinatorial Theory A* **24,** 24–33.

[28] JABBOUR-HATTAB, J. (1999). Martingales et grandes deviations pour les arbres binaires de recherche. *C. R. Acad. Sci. Paris* **328,** 805–810.

[29] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110,** 177–245.

[30] JANSON, S. (2008). Plane recursive trees, Stirling permutations and an urn model. In *Proc. 5th Colloquium Math. Comput. Sci.* (Blaubeuren, 2008), Discrete Mathematics and Theoretical Computer Science Proceedings, Nancy, **AI,** pp. 541–547.

[31] JANSON, S., KUBA, M. AND PANHOLZER, A. (2008). Generalized Stirling permutations, families of increasing trees and urn models. Preprint. Available at http://arxiv.org/abs/0805.4084.

[32] KATONA, Z. AND MÓRI, T. F. (2006). A new class of scale free random graphs. *Statist. Prob. Lett.* **76,** 1587–1593.

[33] KOBER, H. (1952). *Dictionary of Conformal Representations*. Dover Publications, New York.

[34] MCKENZIE, A. AND STEEL, M. (2000). Distributions of cherries for two models of trees. *Math. Biosci.* **164,** 81–92.

[35] MITZENMACHER, M. (2006). A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1,** 226–251.

[36] MÓRI, T. F. (2002). On random trees. *Studia Sci. Math. Hung.* **39,** 143–155.

[37] NAJOCK, D. AND HEYDE, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Prob.* **19,** 675–680.

[38] OLVER, F. W. J. (1974). *Asymptotics and Special Functions*. Academic Press, New York.

[39] ROSENBERG, N. A. (2006). The mean and variance of the numbers of r-pronged nodes and r-caterpillars in Yule-generated genealogical trees. *Ann. Combinatorics* **10,** 129–146.

[40] RUDAS, A., TÓTH, B. AND VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31,** 186–202.

[41] SMYTHE, R. T. AND MAHMOUD, H. M. (1994). A survey of recursive trees. *Teor. Ǐmovīr. Mat. Statist.* 1–29 (in Ukranian). English translation: *Theory Prob. Math. Statist.* **51,** 1–27.

[42] SZYMAŃSKI, J. (1987). On a nonuniform random recursive tree. In *Random Graphs* (Poznań, 1985; 144 of *North-Holland Math. Stud.* **144**), North-Holland, Amsterdam, pp. 297–306.

[43] YULE, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. *Phil. Trans. R. Soc. London B* **213,** 21–87.

[44] ZHANG, J. X. AND DUPUIS, P. (2008). Large-deviation approximations for general occupancy models. *Combin. Prob. Comput.* **17,** 437–470.