A LARGE DEVIATION RATE AND CENTRAL LIMIT THEOREM FOR HORTON RATIOS*

STANLEY XI WANG† AND EDWARD C. WAYMIRE‡

Abstract. Although originating in hydrology, the classical Horton analysis is based on a geometric progression that is widely used in the empirical analysis of branching patterns found in biology, atmospheric science, plant pathology, etc., and more recently in tree register allocation in computer science. The main results of this paper are a large deviation rate and a central limit theorem for Horton bifurcation ratios in a standard network model. The methods are largely self-contained. In particular, derivations of some previously known results of the theory are indicated along the way.

Key words. random tree, bifurcation ratio, large deviation, central limit theorem

AMS(MOS) subject classifications. primary 60F10,60F05,60C05; secondary 86A05,60J85

1. Introduction. River systems around the world are known to hydrologists largely with the aid of maps. This has led to a number of interesting statistical/geometrical observations about rivers, which are understood to varying degrees of empiricism and mathematical rigor. In the present paper we consider a statistic introduced by Horton [19], called the *stream order statistic*, to measure the bifurcation complexity in river networks. Among other things, this statistic has been used to provide an estimate on the total length of rivers in the United States at roughly three million miles; see Leopold [21]. In addition, applications to other naturally occurring branching patterns can be found in Horsfield [17], [18], Berry and Bradley [3], Borchert and Slade [4], Steingraeber, Kascht, and Frank [25], Aho, Sethi, and Ullman [1], Flajolet and Odlyzko [11], Flajolet and Prodinger [10], Flajolet, Raoult, and Vuillemin [12], and Vauchaussade and Viennot [26], to name a few. Some of these and other references can also be found in Jarvis and Woldenberg [20].

For a theoretical formulation of Horton's order analysis, geomorphologists and hydrologists consider an idealized river network represented by a rooted binary tree digraph having n degree-one vertices representing sources; consult Shreve [24] and Chartrand and Lesniak [6] for some of the graph theory terminology. The number n of sources is called the *network magnitude*. Edges of the graph connecting sources to adjacent *junctions* (degree three vertices) are called *external links*, and those between two junctions are called *internal links*. The edge incident to the root is called the *stem*. The stem is regarded as an external link if and only if $n = 1$; otherwise the stem is internal. Each external link is said to have order one. An edge incident to two order-one links is then defined to have order two. We now proceed inductively as follows. An edge has order $k > 1$ if and only if it is incident to two edges of order i and j such that either $i = j = k - 1$ or $i \neq j$ and $\max\{i, j\} = k$. A *stream* (and its order) is defined as a maximal connected path of incident edges of the same order (which

---

is then taken to be the *stream order*). Thus the order-one streams are precisely the (order-one) external links. The maximal order taken over streams in the network is called the *network order*.

Fluctuations in network statistics are represented by assuming all networks of the same magnitude n to be equiprobable, called the *random model*. Because of an inherent imprecision in the small scale details of large river networks, we typically would like robust (large source number) asymptotics for the network variables. Precise asymptotics on the expected (i.e., phase average) network order (maximum stream order) have been obtained by Meir, Moon, and Pounder [22], who show that for fixed n, the expected order is $\frac{1}{2} \log_2 n + O(1)$. Shreve has conjectured on the basis of computer simulation that the location of the mode should also roughly coincide with this value, the details of which are clarified by the results of Meir, Moon, and Pounder [22]. Note that for a given value of n, the largest network order possible is $1 + \log_2 n$. Our focus here is on the asymptotics for sample averages of the lower-order streams and links. Specifically, let $L_{i,n}$ and $S_{i,n}$ denote the sample numbers of links and streams, respectively, of order i in a network of magnitude n. These represent the sample values we compute from a network map. The ratios $L_{2,n}/L_{1,n} = L_{2,n}/n$ and $S_{2,n}/S_{1,n} = S_{2,n}/n$ are called Horton link and stream number *bifurcation ratios*, respectively. Empirical forms of *Horton's laws* refer to the asymptotic stability of these ratios for basins of large magnitude.

Noting a simple mean and variance computation by Werner [27], Gupta and Waymire [15] provide the theoretical counterpart in the form of a law of large numbers for the stream number bifurcation ratios and Mesa [23] for the link numbers; see also Gupta and Waymire [14] for a related overview. The purpose of the present paper is to provide a description of the fluctuations in the form of central limit theorems and large deviation rates where possible. We consider two statistics, the stream number bifurcation ratio and the link number bifurcation ratio. We first obtain a large deviation rate for the former ratio, from which a central limit theorem will also follow. Although we have not obtained the corresponding descriptions of the fluctuations for the link number bifurcation ratio, the computations given in § 4 suggest that similar results should hold for this ratio.

Some very interesting results have already appeared in the literature, which provide Gaussian asymptotic approximations to combinatorial enumerations, e.g., see Carlitz et al. [5], Harper [16], Flajolet and Odlyzko [11], and Bender [2]. In particular, Bender [2] provides a general condition for the asymptotic normality of a doubly indexed sequence of positive numbers that essentially requires a pole in the bivariate generating function. Bender [2] also gives a number of interesting example applications. However, these results do not seem to be applicable to the present problem since the singularity in the bivariate generating function is not a pole; see the remark at the end of the next section.

The precise statements of main results are given in § 2. Various other results are obtained along the way, including the laws of large numbers, which serve to unify the previously known asymptotics and exact formulae for stream and link number probabilities and their expected values. In addition, a few new exact formulae are also provided in this connection. The proofs of the main results are given in § 5. Both the calculation of large deviation rates and the central limit theorem rest on the natural recursive structure of random model described in § 3. An analysis of the asymptotic form of the factorial moments of the distribution of stream and link numbers is given in § 4, which may be read independently of the proofs of the main results. In particular, we obtain the correct *factorial moments* of a Poisson distribution with parameter

$\lambda \propto n$ for the asymptotic (in n) factorial moments of the bifurcation ratios, but the (correct) asymptotic Gaussian approximation to this Poisson distribution is wrong for the bifurcation ratios (i.e., right mean, wrong variance parameter). These factorial moment calculations illustrate the delicacy of the problems due to the occurence of $\infty - \infty$ effects.

2. Statement of results. Let Ω_n denote the collection of rooted binary tree graphs of magnitude n. Then, according to a classic (1858) formula of Cayley, its cardinality $|\Omega_n|$ is given by

$$(2.1) \qquad |\Omega_n| = \frac{1}{2n-1} \binom{2n-1}{n} \sim \frac{2^{2n-2}}{\sqrt{\pi}} n^{-3/2}, \quad n \to \infty.$$

The random model is defined by the probability measure P_n, which assigns probability $|\Omega_n|^{-1}$ to each $\tau \in \Omega_n$. The random variables $L_{i,n}, S_{i,n}$, denoting link numbers and stream numbers of order i, respectively, are defined on Ω_n as in the Introduction. The random variables $R_{L,2}^{(n)} = L_{2,n}/n$ and $R_{S,2}^{(n)} = S_{2,n}/n$ are referred to as *link* and *stream number bifurcation ratios*, respectively.

It is now well known that the value of each of the bifurcation ratios $S_{2,n}/n$ and $L_{2,n}/n$ stabilizes according to the following law of large numbers.

THEOREM 2.1. (Law of large numbers). *For the random model,* (i)$S_{2,n}/n \to \frac{1}{4}$ *in probability as* $n \to \infty$; (ii) $L_{2,n}/n \to \frac{1}{2}$ *in probability as* $n \to \infty$.

In fact, this may be obtained by the methods of the present paper according to which one has the following properties.

PROPOSITION 2.2. *For* $n \geq 3$, (i)

$$E\{S_{2,n}\} = \frac{n(n-1)}{2(2n-3)} \sim \frac{n}{4}$$

$$E\{S_{2,n}(S_{2,n}-1)\} = \frac{n(n-1)(n-2)(n-3)}{4(2n-5)(2n-3)} \sim \frac{n^2}{16};$$

(ii)

$$E\{L_{2,n}\} = \frac{2n-1}{\binom{2n-1}{n}} \left\{ 2^{2n-3} - 2^{n-2} - \sum_{m=1}^{n-2} \binom{2m}{m} \frac{2^{n-m-2}}{2m-1} (2^{n-m-1} - 1) \right\} \sim \frac{n}{2}$$

$$E\{L_{2,n}(L_{2,n}-1)\} = \frac{2n-1}{\binom{2n-1}{n}} \left\{ 4(2n-5) \binom{2n-6}{n-3} \right.$$
$$\left. - \sum_{k=0}^{n-4} (3n - 3k - 13)(2k+1) \binom{2k}{k} 2^{n-k-3} \right\} \sim \frac{n^2}{4}.$$

The exact expressions in (ii) are new, but the exact forms in (i) have been obtained previously by other methods. It is important to note from the exact calculations that the asymptotic formulae for the first two factorial moments do *not* provide the asymptotic variance; i.e., there is an $\infty - \infty$ contribution. For example,

$$\begin{aligned} (2.2) \qquad \mathrm{Var} S_{2,n} &= E S_{2,n}(S_{2,n}-1) + E S_{2,n} - (E S_{2,n})^2 \\ &= \frac{n(n-1)(n-2)(n-3)}{2(2n-3)(2n-3)(2n-5)} \sim \frac{n}{16} \qquad , \end{aligned}$$

while

(2.3)
$$\frac{n^2}{16} + \frac{n}{4} - (\frac{n}{4})^2 = \frac{n}{4}.$$

This point is significant in a theory that rests largely on asymptotics, as is further demonstrated in §4. This aside, note that Theorem 2.1 follows from the linearity of the mean and variance in n (Proposition 2.2) by an application of Chebyshev's inequality.

Proposition 2.2 is proved in § 4 together with a generalization to the precise asymptotic forms of the higher-order factorial moments. The moment analysis is based on the following identity which may be of independent interest.

LEMMA 2.1. *Let $\Gamma(x)$ denote the gamma function. Then, for each $n = 1, 2, \cdots$, we have*

(2.4)
$$\sum_{i=1}^{n} \binom{n+1}{i} \Gamma(i - \frac{1}{2})\Gamma(n - i + \frac{1}{2}) = 4\sqrt{\pi}\Gamma(n + \frac{1}{2}).$$

The identity (2.4) has several interesting variants. For example,

(2.5)
$$\sum_{i=1}^{n} \binom{n+1}{i} \binom{n-1}{i-1} \binom{2n-2}{2i-2}^{-1} = 2(2n - 1).$$

Remark. This lemma is used in the moment computations given in § 4. Its verification is quite amenable to "proof" by symbolic algebra software; e.g., Macsyma or Maple. This was, in fact, our first approach (after calculator tests) to checking the assertion. However, the result can also be obtained as a case of Gauss's theorem for $_2F_1$ hypergeometric functions or by induction, as indicated below. For us, the identity (2.4) was uncovered by the process of "matching asymptotics," i.e., in trying to identify the slowly varying part of the asymptotic Tauberian expansion (4.11) below.

Proof of Lemma 2.1.

Method 1 (Classical identities). We have

(2.6)
$$_2F_1 \begin{pmatrix} a, b \\ \\ c \end{pmatrix} ; 1 \end{pmatrix} = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)}.$$

From this we note in the cases where $i = 0$ and $i = n + 1$ that

(2.7)
$$\binom{n+1}{i} \Gamma(i - \frac{1}{2})\Gamma(n - i + \frac{1}{2}) = -2\sqrt{\pi}\Gamma(n + \frac{1}{2}),$$

so that the identity of Lemma 2.1 is reduced to

(2.8)
$$\Gamma(-\frac{1}{2})\Gamma(n + \frac{1}{2}) \,_2F_1 \begin{pmatrix} -n-1 & , & -\frac{1}{2} \\ \\ & \frac{1}{2} - n \end{pmatrix} ; 1 \end{pmatrix} = 0.$$

Likewise, the equivalent variant (2.5) may be obained as a specialization of the (Hagan/Rothe) identity (3.146) given in Gould [13]; take $y = 2n + 1, p = n + 1, x = 1, q = -1$, and $z = 2$, there.

Method 2 (Induction). Let $B(u, v)$ denote a Beta function. It is well known that

$$B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u + v)}.$$

Thus, it suffices to show that

$$(2.9) \qquad \frac{1}{2\sqrt{\pi}}(n - 1)! \sum_{i=1}^{n} \binom{n + 1}{i} \int_0^{\frac{\pi}{2}} (\cos\theta)^{2(i-1)}(\sin\theta)^{2(n-i)} d\theta = \Gamma(n + \frac{1}{2}).$$

By induction, suppose for $k \leq n$ (2.9) is true; then for $k = n + 1$, we have

$$\text{lhs} = \frac{1}{2\sqrt{\pi}} n! \sum_{i=1}^{n+1} \binom{n + 2}{i} \int_0^{\pi/2} (\cos\theta)^{2(i-1)}(\sin\theta)^{2(n+1-i)} d\theta$$

$$= \frac{1}{2\sqrt{\pi}} n! \sum_{i=1}^{n} [\binom{n + 1}{i} + \binom{n + 1}{i - 1}] \int_0^{\pi/2} (\cos\theta)^{2(i-1)}(\sin\theta)^{2(n+1-i)} d\theta$$

$$+ \frac{1}{2\sqrt{\pi}} n! \binom{n + 2}{n + 1} \int_0^{\pi/2} (\cos\theta)^{2n} d\theta$$

$$= n\Gamma(n + \frac{1}{2}) - \frac{1}{2\sqrt{\pi}} n \sum_{i=1}^{n} (n - 1)! \binom{n + 1}{i} + \int_0^{\pi/2} (\cos\theta)^{2i}(\sin\theta)^{2(n-i)} d\theta$$

$$+ \frac{n!}{2\sqrt{\pi}} \sum_{i=0}^{n-1} \binom{n + 1}{i} + \int_0^{\pi/2} (\cos\theta)^{2i}(\sin\theta)^{2(n-i)} d\theta + \frac{n!(n + 2)}{2\sqrt{\pi}} \int_0^{\pi/2} (\cos\theta)^{2n} d\theta$$

$$= n\Gamma(n + \frac{1}{2}) + \frac{n!}{2\sqrt{\pi}} \frac{\Gamma(n + \frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(n + 1)} = \Gamma(n + \frac{3}{2}) = \text{rhs}.$$

To accompany the law of large numbers it is important to have some measure of the fluctuations from the average. The idea behind the large deviation rate is that the probability of a deviation from the mean by some prescribed amount goes to zero at an exponentially fast rate, which we may try to calculate. The following results describe probabilities of fluctuations from this point of view.

THEOREM 2.3. (Large deviation rate). *For the random model,*

$$(2.10) \qquad \lim_{n\to\infty} \frac{1}{n}\log P(\frac{S_{2,n}}{n} > y) = -I(y), y \in (\frac{1}{4}, \frac{1}{2})$$

and

$$(2.11) \qquad \lim_{n\to\infty} \frac{1}{n}\log P(\frac{S_{2,n}}{n} < y) = -I(y), y \in (0, \frac{1}{4}),$$

where

$$(2.12) \qquad I(y) = (4y - 1)\tanh^{-1}(4y - 1) - \log(\cosh(\tanh^{-1}(4y - 1))).$$

The rate function $I(y)$ is called an *entropy function* in the theory of large deviations; see Ellis [9]. The graph of the entropy function (2.12) is a ∪-shape on the interval $(0, \frac{1}{2})$ with a minimum at $(\frac{1}{4}, 0)$.

THEOREM 2.4. (Central Limit Theorem). *For the random model, we have*

$$(2.13) \qquad \sqrt{n}\left(\frac{S_{2,n}}{n} - \frac{1}{4}\right) \Rightarrow N\left(0, \frac{1}{16}\right), \quad n \to \infty,$$

where $\Rightarrow$ *denotes convergence in distribution, and* $N(\mu, \sigma^2)$ *denotes a limit law that is Gaussian having mean* μ *and variance* σ^2.

Remark. Let $l_i(n,k), s_i(n,k)$ denote the number of trees in Ω_n having k links of order i and k streams of order i, respectively, as defined in the Introduction, and define

$$(2.14) \qquad \hat{l}_2(x,y) = \sum_{n,k} l_2(n,k) x^n y^k,$$

$$(2.15) \qquad \hat{s}_2(x,y) = \sum_{n,k} s_2(n,k) x^n y^k.$$

Also, let

$$(2.16) \qquad \hat{l}_2'(x,y) = \sum_{n,k} l_2'(n,k) x^n y^k,$$

where $l_2'(n,k)$ denotes the number of trees of network order 2 in Ω_n having k links of order 2. Then considerations of the recursive structure give the following relations:

$$(2.17) \qquad \hat{s}_2 = \hat{s}_2^2 + 2x\hat{s}_2 + x^2 y,$$

$$(2.18) \qquad \hat{l}_2' = x^2 y + 2xy\hat{l}_2',$$

$$(2.19) \qquad \hat{l}_2 = \hat{l}_2' + 2x(\hat{l}_2 - \hat{l}_2') + \hat{l}_2^2.$$

Solving for these generating functions, we find singularities other than poles. This does not seem to be covered by general theory; cf. Bender [2].

3. Some preliminaries. We continue to let $l_i(n,k), s_i(n,k)$ denote the number of trees in Ω_n having k links of order i and k streams of order i, respectively. Then for the random model

$$(3.1) \qquad P_n(L_{i,n} = k) = \frac{l_i(n,k)}{|\Omega_n|},$$

$$(3.2) \qquad P_n(S_{i,n} = k) = \frac{s_i(n,k)}{|\Omega_n|}, \quad n \geq 1, k \geq 0.$$

In view of the recursive structure of the trees as described precisely in Meir, Moon, and Pounder [22], we obtain convolution identities among the $s_2(n,k)$'s and $l_2(n,k)$'s of the following forms.

LEMMA 3.1. *For the random model,*

(3.3a)   (i) $s_2(1,k) = \delta_{0,k}, \; s_2(2,k) = \delta_{1,k}, \; s_2(4,1) = 4, \; s_2(4,2) = 1,$

(3.3b) $s_2(n,0) = 0, \quad s_2(n,1) = 2^{n-2}, \; n \geq 2,$

(3.3c) $s_2(n,k) = \displaystyle\sum_{m=1}^{n-1}\sum_{j=0}^{k} s_2(m,j)s_2(n-m,k-j), \; n \geq 5, k \neq 2;$

(3.4a)   (ii) $l_2(1,k) = \delta_{0,k}, \; l_2(2,k) = \delta_{1,k}, \; l_2(3,k) = 2\delta_{2,k}, l_2(n,0) = 0,$

(3.4b) $l_2(n,n-1) = 2^{n-2}, l_2(n,n-2) = \displaystyle\sum_{m=2}^{n-2} l_2(m,m-1)l_2(n-m,n-m-1),$

(3.4c) $l_2(n,k) = \displaystyle\sum_{m=1}^{n-1}\sum_{j=0}^{k} l_2(m,j)s_2(n-m,k-j), \; n \geq 5, k \leq n-3.$

The proof of these recursions are fairly straightforward and will be left to the reader. It is to be noted that, in either case, special provision must be made for the *order-three networks* (i.e., $n = 4$, $k = 2$ for stream count, $n = 4, k = n - 2$ for link count).

The convolution forms in the identities of Lemma 3.1 transform under

(3.5) $\hat{s}_2(n,x) = \displaystyle\sum_{k=1}^{n} s_2(n,k)x^k,$

(3.6) $\hat{l}_2(n,x) = \displaystyle\sum_{k=0}^{n} l_2(n,k)x^k,$

according to the following lemma.

LEMMA 3.2.

(3.7a)   (i) $\hat{s}_2(1,x) = 1, \hat{s}_2(2,x) = x, \hat{s}_2(3,x) = 2x, \hat{s}_2(4,x) = x^2 + 4x,$

(3.7b) $\hat{s}_2(n,x) = \displaystyle\sum_{m=1}^{n-1} \hat{s}_2(m,x)\hat{s}_2(n-m,x), n \geq 3,$

and

(3.8a)   (ii) $\hat{l}_2(1,x) = 1, \; \hat{l}_2(2,x) = x,$

$$(3.9b) \qquad \hat{l}_2(n, x) = (2x)^{n-2}(x-1) + \sum_{m=1}^{n-1} \hat{l}_2(m, x)\hat{l}_2(n-m, x), n \geq 3.$$

Although somewhat lengthy, Lemma 3.2 is easily verified as a consequence of Lemma 3.1 and its proof is omitted; the details for (ii) can be found in Mesa [23]. These recursions represent the starting point for our proofs. Define moment generating functions

$$(3.10) \qquad \psi_n(\xi) = Ee^{\xi S_{2,n}} = \sum_k e^{k\xi} P_n(S_{2,n} = k) = |\Omega_n|^{-1}\hat{s}_2(n, e^\xi)$$

and

$$(3.11) \qquad \lambda_n(\xi) = Ee^{\xi L_{2,n}} = \sum_k e^{k\xi} P_n(L_{2,n} = k) = |\Omega_n|^{-1}\hat{l}_2(n, e^\xi).$$

Then Lemma 3.2 provides recursive equations of convolution type for $|\Omega_n|\psi_n(\xi)$ and $|\Omega_n|\lambda_n(\xi)$ which will be analyzed in §§4 and 5.

For ease in reference, we close this section with the statements of the theorem to be used in the proofs in § 5. Theorem 3.3 seems to have a somewhat fragmented history and has been useful in diverse contexts; see Cox and Griffeath [7] and references therein. A systematic treatment of the elements of large deviation theory can be found in Ellis [9] and Deuschel and Stroock [8].

THEOREM 3.3. (Sievers,Plachky and Steinbach, Ellis, Cox and Griffeath). *Let $\{X_n : n = 0, 1 \cdots\}$ be a sequence of random variables and let*

$$(3.12) \qquad \varphi_n(\xi) = a_n^{-1}\log Ee^{\xi X_n},$$

where $\{a_n\}$ is a sequence of positive numbers such that $a_n \to \infty$ as $n \to \infty$. Assume that on the interval $(\xi_-, \xi_+) \ni 0$, we have

$$(3.13) \qquad \lim_{n\to\infty} \varphi_n(\xi) = \varphi_\infty(\xi) < \infty,$$

where $\varphi_\infty(\xi)$ is strictly convex and C^2 on (ξ_-, ξ_+). If φ_n' is convex on $[0, \xi_+)$, and $\lim_{n\to\infty} \varphi_n''(0) = \sigma^2 = \varphi_\infty''(0)$, then

$$(3.14) \qquad \lim_{n\to\infty} a_n^{-1}\log P(\frac{X_n}{a_n} > y) = -I(y), y \in (\mu, \alpha_+)$$

and

$$(3.15) \qquad \lim_{n\to\infty} a_n^{-1}\log P(\frac{X_n}{a_n} < y) = -I(y), y \in (\alpha_-, \mu),$$

where $\mu = \varphi_\infty'(0), \alpha_- = \varphi_\infty'(\xi_-+), \alpha_+ = \varphi_\infty'(\xi_+-)$, and $I(y)$ is the Legendre transform of $\varphi_\infty(\xi)$. In addition,

$$(3.16) \qquad \frac{X_n - EX_n}{\sqrt{a_n}} \Rightarrow N(0, \sigma^2), \quad n \to \infty.$$

4. A moment analysis. This section may be read independently of the proofs of the main results of the paper given in the next section.

The factorial moments are defined by

$$(4.1) \qquad \nu_n^{[r]} = E_n[S_{2,n}]_r = |\Omega_n|^{-1} \frac{d^r}{dx^r} \hat{s}_2(n,x)|_{x=1},$$

$$(4.2) \qquad \mu_n^{[r]} = E_n[L_{2,n}]_r = |\Omega_n|^{-1} \frac{d^r}{dx^r} \hat{l}_2(n,x)|_{x=1},$$

where $[x]_r = x(x-1)\cdots(x-r+1)$, $x \in R$, $r = 1, 2 \cdots$. In view of Lemma 3.2, we have accordingly, for $\overline{\nu}_n^{[r]} = |\Omega_n|\nu_n^{[r]}$ and $\overline{\mu}_n^{[r]} = |\Omega_n|\mu_n^{[r]}$, that for each $r = 1, 2, \cdots, n \geq 4$,

$$(4.3) \qquad \overline{\nu}_n^{[r]} = \sum_{m=1}^{n-1} \sum_{j=0}^{r} \binom{r}{j} \overline{\nu}_n^{[r-j]} \overline{\nu}_{n-m}^{[j]}$$

(4.4)
$$\overline{\mu}_n^{[r]} = \left\{ \frac{1}{n-1} \sum_{m-1}^{n} \delta_{r,m} \right\} r(n-1)(n-2)\cdots(n-r)2^{n-2} + \sum_{m=1}^{n-1} \sum_{j=0}^{r} \binom{r}{j} \overline{\mu}_n^{[r-j]} \overline{\mu}_{n-m}^{[j]}.$$

To obtain these, simply note that

$$\frac{d^r}{dx^r} \hat{s}_2(n,x) = \sum_{m=1}^{n-1} \sum_{j=0}^{r} \binom{r}{j} \frac{d^{r-j}}{dx^{r-j}} \hat{s}_2(m,x) \frac{d^j}{dx^j} \hat{s}_2(n-m,x),$$

$$\frac{d^r}{dx^r} \hat{l}_2(n,x) = \left\{ 2^{n-2}(n-1)(n-2)\cdots(n-r)[(n-1)x - n + r + 1]x^{n-r-2} \right\} \frac{1}{n-1} \sum_{m-1}^{n} \delta_{r,m}$$

$$+ \sum_{m=1}^{n-1} \sum_{j=0}^{r} \binom{r}{j} \frac{d^{r-j}}{dx^{r-j}} \hat{l}_2(m,x) \frac{d^j}{dx^j} \hat{l}_2(n-m,x),$$

Proof of Proposition. 2.2. To verify (ii), for example, use (4.4) with $\overline{\mu}_1^{[1]} = 0$ and $\overline{\mu}_2^{[1]} = 1$. Then, $\hat{\mu}_1(t) = \sum_{n=1}^{\infty} \overline{\mu}_n^{[1]} t^n$ satisfies

$$\hat{\mu}_1(t) = t^2 + \sum_{n=3}^{\infty} 2^{n-2} t^n + 2 \sum_{n=3}^{\infty} \sum_{m=1}^{n-1} \overline{\mu}_m^{[1]} |\Omega_{n-m}| t^n$$

$$= \frac{t^2}{1-2t} + 2\hat{\mu}_1(t)\hat{\Omega}(t),$$

where $\hat{\Omega}(t) = \sum_{n=1}^{\infty} |\Omega_n| t^n = (1 - \sqrt{1-4t})/2$. Thus,

$$\hat{\mu}_1(t) = t^2 \left\{ \frac{-1}{1-2t} + \frac{2}{1-4t} \right\} (1-4t)^{1/2}$$

$$= \sum_{n=0}^{\infty} \{2 \cdot 4^n - 2^n\} t^{n+2} \sum_{m=0}^{\infty} \binom{\frac{1}{2}}{m} (-4t)^m$$

$$= \sum_{n=2}^{\infty} \left\{ \sum_{\{(m,k):m+k=n,k\geq 2,m\geq 0\}} \{2 \cdot 4^{k-2} - 2^{k-2}\} \binom{\frac{1}{2}}{m} (-4)^m \right\} t^n.$$

Thus,

$$(4.5) \qquad \overline{\mu}_n^{[1]} = \sum_{m=0}^{n-2} \{2 \cdot 4^{n-m-2} - 2^{n-m-2}\} \binom{\frac{1}{2}}{m} (-4)^m, n \geq 2.$$

Since

$$\binom{\frac{1}{2}}{m} = 2^{-2m}(-1)^{m-1}\frac{1}{2m-1}\binom{2m}{m}, m \geq 1,$$

we have

$$(4.6) \quad \overline{\mu}_n^{[1]} = 2 \cdot 4^{n-2} - 2^{n-2} + \sum_{m=1}^{n-2}(-1)^{2m-1}\frac{1}{2m-1}\binom{2m}{m}\{2^{2n-2m-3} - 2^{n-m-2}\}.$$

Also, $\overline{\mu}_1^{[2]} = 0$ and $\overline{\mu}_2^{[2]} = 0$, Then, $\hat{\mu}_2(t) = \sum_{n=1}^{\infty} \overline{\mu}_n^{[2]} t^n$ satisfies

$$\hat{\mu}_2(t) = \sum_{n=3}^{\infty}\sum_{m=1}^{n-1}\sum_{j=0}^{2}\binom{2}{j}\overline{\mu}_m^{[2-j]}\overline{\mu}_{n-m}^{[j]}t^n + 2\sum_{n=3}^{\infty}(n-2)2^{n-2}t^n$$

$$= \frac{4t^3}{(1-2t)^2} + 2\sum_{n=3}^{\infty}\sum_{m=1}^{n-1}\overline{\mu}_{n-m}^{[2]}|\Omega_m|t^n + 2\sum_{n=3}^{\infty}\sum_{m=1}^{n-1}\overline{\mu}_m^{[1]}\overline{\mu}_{n-m}^{[1]}t^n$$

$$= \frac{4t^3}{(1-2t)^2} + 2\hat{\mu}_2(t)\hat{\Omega}(t) + 2\hat{\mu}_1^2(t)$$

$$(4.7) \qquad = \frac{4t^3 - 14t^4}{(1-4t)^{\frac{3}{2}}(1-2t)^2}.$$

On the other hand,

$$\sum_{j=0}^{\infty}\overline{\mu}_{j+3}^{[2]}t^j = \frac{\hat{\mu}_2(t)}{t^3}$$

$$= 4(1-4t)^{-3/2}(1-2t)^{-2} - 14t(1-4t)^{-3/2}(1-2t)^{-2}$$

$$= \sum_{k=0}^{\infty}\binom{\frac{-3}{2}}{k}(-4)^k t^k \sum_{m=1}^{\infty}m(2t)^{m-1} - 14t\sum_{k=0}^{\infty}\binom{\frac{-3}{2}}{k}(-4)^k t^k \sum_{m=1}^{\infty}m(2t)^{m-1}$$

$$= 4\sum_{n=0}^{\infty}\sum_{m+k=n}^{\infty}\binom{\frac{-3}{2}}{k}(-4)^k 2^{n-k}(n-k+1)t^n$$

$$\quad - 14\sum_{n=0}^{\infty}\sum_{m+k=n}^{\infty}\binom{\frac{-3}{2}}{k}(-4)^k 2^{n-k}(n-k+1)t^{n+1}$$

$$= 4 + \sum_{j=1}^{\infty}\{4\binom{\frac{-3}{2}}{j}(-4)^j + \sum_{k=0}^{j-1}\binom{\frac{-3}{2}}{k}(-4)^k 2^{j-k}(3k-3j+4)\}t^j.$$

Thus, $\overline{\mu}_3^{[2]} = 4$, and

$$\overline{\mu}_{j+3}^{[2]} = 4\binom{\frac{-3}{2}}{j}(-4)^j + \sum_{k=0}^{j-1}\binom{\frac{-3}{2}}{k}(-4)^k 2^{j-k}(3(k-j)+4), \qquad j \geq 1,$$

i.e.,

$$\overline{\mu}_n^{[2]} = 4 \binom{\frac{-3}{2}}{n-3} (-4)^{n-3} + \sum_{k=0}^{n-4} \binom{\frac{-3}{2}}{k} (-4)^k 2^{n-k-3} (3k - 3n + 13), n \geq 4.$$

Since

$$\binom{\frac{-3}{2}}{k} = 2^{-2k} (-1)^k (2k+1) \binom{2k}{k},$$

we obtain,

$$E\{L_{2,n}(L_{2,n} - 1)\} = 4(2n-5) \binom{2n-6}{n-3}$$

(4.8)
$$- \sum_{k=0}^{n-4} (3n - 3k - 13)(2k+1) \binom{2k}{k} 2^{n-k-3}.$$

The proof of part (i) of Proposition 2.2 is similar and left to the reader.

For the higher-order moments, the Tauberian theorem can be used with Lemma 2.1 to show for each $r \geq 1$,

(4.9)
$$E[S_{2,n}]_r \sim (\frac{n}{4})^r, \qquad n \to \infty$$

(4.10)
$$E[L_{2,n}]_r \sim (\frac{n}{2})^r, \qquad n \to \infty.$$

Following is how we obtain (4.10) from Lemma 2.1 and the above. The case of (4.9) is similar. When the Tauberian theorem is applied, we need only consider the terms having highest power of $(1 - 4t)$ in the dominators of (4.4) and, by an induction argument applied to (4.4), we may check that $\hat{\mu}_r(t) = \sum_{n=1}^{\infty} \overline{\mu}_n^{[r]} t^n$ is $O((1 - 4t)^{-(2r-1)/2})$ as $t \to \frac{1}{4}$. With this observation, we are ready to show (4.10). Again by induction, suppose that (4.10) is true for $r \leq k$; so by the Tauberian theorem, we have

(4.11)
$$\overline{\mu}_n^{[k]} \sim \frac{1}{\Gamma(k - \frac{1}{2})} n^{k-(3/2)} \Lambda_k(n) 2^{2n},$$

where the last term on the right come from a change of variable of the form $u = 4t$.

Since $\overline{\mu}_n^{[r]} = |\Omega_n| \mu_n^{[r]}$, and by (2.1), we have

(4.12)
$$\Lambda_k(n) = \frac{1}{4\sqrt{\pi}} \Gamma(k - \frac{1}{2})(\frac{1}{2})^k.$$

Now for $r = k + 1$, we have

(4.13)
$$\overline{\mu}_n^{[k+1]} \sim \frac{1}{\Gamma(k + \frac{1}{2})} k^{n-1/2} \Lambda_{k+1}(n) 2^{2n}.$$

By (4.4), we have

(4.14)
$$\Lambda_{k+1}(n) = \sum_{j=0}^{k+1} \binom{k+1}{j} \Lambda_{k+1-j}(n) \Lambda_j(n).$$

By (4.12) and Lemma 2.1, (4.14) is readily simplified. Plug this into (4.13) and the result follows.

5. Proof of main result. Consider first $\psi_n(\xi)$. Note that

$$(5.1a) \qquad \psi_1(\xi) = 1, \qquad \psi_2(\xi) = \psi_3(\xi) = e^\xi.$$

We will show that

$$(5.1b) \quad \psi_n(\xi) = \frac{(2n-1)}{\binom{2n-1}{n}} \sum_{k=0}^{[\frac{n}{2}]} \binom{n-k}{k} \frac{1}{2(n-k)-1} \binom{2n-2k-1}{n-k} (e^\xi - 1)^k.$$

Let $a_n(\xi) = |\Omega_n|\psi_n(\xi)$; then $a_1(\xi) = 1, a_2(\xi) = e^\xi, a_3(\xi) = 2e^\xi$. Also, by (3.7b),

$$a_n(\xi) = \sum_{m=1}^{n-1} a_m(\xi)a_{n-m}(\xi).$$

Then, $\hat{a}(s,\xi) = \sum_{n=1}^\infty a_n(\xi)s^n$ satisfies

$$\hat{a}(s,\xi) = s + s^2(e^\xi - 1) + \sum_{m=1}^\infty \sum_{n=m+1}^\infty a_m(\xi)s^m a_{n-m}(\xi)s^{n-m}.$$

Thus,

$$\hat{a}(s,\xi) = \frac{1}{2}\{1 - \sqrt{1 - 4s - 4s^2(e^\xi - 1)}\}$$

$$= \frac{1}{2} \sum_{n=1}^\infty \binom{\frac{1}{2}}{n} (-1)^{n+1}(4s)^n(1 + s(e^\xi - 1))^n$$

$$= \frac{1}{2} \sum_{n=1}^\infty \sum_{k=0}^\infty \binom{\frac{1}{2}}{n} (-1)^{n+1}(4)^n \binom{n}{k} (e^\xi - 1)^k s^{n+k}$$

$$(5.2) \qquad = \frac{1}{2} \sum_{m=1}^\infty \sum_{\{(n,k):n+k=m,n\geq k \wedge 1\}} \binom{\frac{1}{2}}{n} (-1)^{n+1}(4)^n \binom{n}{k} (e^\xi - 1)^k s^m.$$

Then we have

$$(5.3) \qquad a_m(\xi) = \frac{1}{2} \sum_{\{(n,k):n+k=m,n\geq k \wedge 1\}} \binom{\frac{1}{2}}{n} \binom{n}{k} (-1)^{n+1}(4)^n (e^\xi - 1)^k.$$

Thus,

$$\psi_n(\xi) = \frac{1}{2|\Omega_n|} \sum_{\{(j,k):j+k=n,j\geq k \wedge 1\}} \binom{\frac{1}{2}}{j} \binom{j}{k} (-1)^{j+1}(4)^j (e^\xi - 1)^k.$$

Note that

$$\binom{\frac{1}{2}}{j} = \frac{2^{-2j+1}(-1)^{j-1}(2j-2)!(2j-1)}{j!(j-1)!(2j-1)}.$$

Thus,

$$(5.4) \qquad \psi_n(\xi) = \frac{1}{|\Omega_n|} \sum_{\{(j,k):j+k=n,j\geq k\wedge 1\}} \binom{j}{k} \frac{1}{2j-1} \binom{2j-1}{j} (e^\xi - 1)^k.$$

By change of variable again, we get (5.1b).

Now using (2.1) and (5.1b), we have

$$(5.5) \qquad \psi_n(\xi) = \sum_{k=0}^{[n/2]} \frac{(n-k)!|\Omega_{n-k}|(e^\xi - 1)^k}{(n-2k)!|\Omega_n|k!},$$

with

$$(5.6) \qquad \frac{|\Omega_{n-k}|}{|\Omega_n|} \sim 4^{-k}(1 - \frac{k}{n})^{-3/2}, \quad k \leq \frac{n}{2}, n \to \infty.$$

In particular, (5.5) may be expressed as

$$(5.7) \qquad \psi_n(\xi) = \sum_{k=0}^{[n/2]} \frac{(1 - \frac{k}{n})^{-\frac{3}{2}}(1 - \frac{k}{n})\cdots(1 - \frac{(2k-1)}{n})(\frac{n}{4}e^\xi - 1)^k}{k!}.$$

The appropriate choice of scaling and the computation of the asymptotic variance as

$$(5.8) \qquad a_n = n \quad \text{and} \quad \sigma^2 = \varphi''_\infty(0)$$

can be determined very simply from (5.7) using (2.2). In particular,

$$(5.9) \qquad \sigma^2 = \lim_{n\to\infty} \varphi''_\infty(0) = \frac{1}{16}.$$

Taking $a_n = n$ in the application of Theorem 3.3, we may check that

$$(5.10) \qquad \lim_{n\to\infty} \varphi_n(\xi) = \frac{\xi}{4} + \log(\cosh(\frac{\xi}{4})) = \varphi_\infty(\xi).$$

This computation in a neighborhood of zero can be made by a saddle point method. In particular, for $\xi \geq 0$ we need only to calculate the maximum term of the sum for k ranging between 0 and $\frac{n}{2}$, and for $\xi < 0$ we calculate the maximum difference between pairwise successive terms $2k$ and $2k + 1$ with k ranging from 1 to $\frac{n}{4}$ (which nicely factors). The parameters $\alpha_- = 0$ and $\alpha_+ = \frac{1}{2}$ and the computation of the Legendre transform as

$$(5.11) \qquad I(y) = (4y - 1)\tanh^{-1}(4y - 1) - \log(\cosh(\tanh^{-1}(4y - 1)))$$

follow. We now apply Theorem 3.3 to get both the large deviation probabilities and the central limit theorem.

REFERENCES

[1] A.V. AHO, R. SETHI, AND J. D. ULLMAN, *Compilers*, Addison–Wesley, Reading, MA, 1986.

[2] E. A. BENDER, *Central and local limit theorems applied to asymptotic enumeration*, J. Combin. Theory (A), 15 (1973), pp. 91–111.

[3] M. BERRY AND P. M. BRADLEY, *The application of network analysis to the study of brannching patterns of large dendritic fields*, Brain Research, 109 (1976), pp. 111–132.

[4] R. BORCHERT AND N. A. SLADE, *Bifurcation ratios and adaptive geometry of trees*, Bot. Gaz.,142 (1981), pp. 394–401.

[5] L. CARLITZ, D. C. KURTZ, R. SCOVILLE, AND O. P. STACKELBERG, *Asymptotic properties of Eulerian numbers*, Z. Wahr. Verw. Geb., 23(1972), pp. 47–54.

[6] G. CHARTRAND AND L. LESNIAK, *Graphs and DiGraphs*, Second edition, Wadsworth, Monterey, CA, 1986.

[7] T. COX AND D. GRIFFEATH, *Large deviations for Poisson systems of independent random walks*, Z. Wahr. Verw. Geb., 66 (1984), pp. 543–558.

[8] J.D. DEUSCHEL AND D. W. STROOCK, *Large Deviations*, Academic Press, Boston, 1989.

[9] R. S. ELLIS, *Entropy, Large Deviations, and Statistical Mechanics* Springer-Verlag, Berlin, 1985.

[10] P. FLAJOLET AND H. PRODINGER, *Register allocation for unary-binary trees*, SIAM J. Comput., 13 (1984), pp. 629–640.

[11] P. FLAJOLET AND A.M. ODLYZKO, *Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations*, Math. Proc. Cambridge Philos. Soc., 96 (1984), pp. 237–253.

[12] P. FLAJOLET, J. C. RAOULT, AND J. VUILLEMIN, *The number of registers recquired for evaluating arithmetic expressions*, Theoret. Comput. Sci., 9 (1979), pp. 99–125.

[13] H. W. GOULD, *Combinatorial Identities*, Morgantown Press, Morgantown, WV, 1972.

[14] V. K. GUPTA AND E. WAYMIRE, *The spatial geometry of random networks and a problem in river basin hydrology*, AMS-IMS-SIAM Conference on Spatial Statistics and Imaging, 1990.

[15] ———, *On the formulation of an analytical approach to hydrologic response and similarity at the basin scale*, J. Hydrol., 65 (1983) pp. 95–123.

[16] L.H. HARPER, *Stirling behavior is asymptotically normal*, Ann. Math. Stat., 38 (1967) pp. 410–414.

[17] K. HORSFIELD, *Are diameter, length, and branching ratios meaningful in the lung?*, J. Theoret. Biol., 87 (1980), pp. 773–784.

[18] ———, *The structure of the tracheobronchial tree*, in Scientific Foundations of Respitory Medicine, J. G. Scadding, G. Cumming, and W. M. Thurlbeck, eds., Heinemann, London, 1981, pp. 54–77.

[19] R. HORTON, *Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology*, Bull. Geol. Soc. Amer., 56, (1945), pp. 275–370.

[20] R.S. JARVIS AND M.J. WOLDENBERG, *River Networks* Benchmark Papers in Geology, 80, Hutchinson Ross, Stroudsburg, PA, 1984.

[21] L. B. LEOPOLD, *Rivers*, American Scientist, 50 (1962), pp. 511–537.

[22] A. MEIR, J. W. MOON, AND J. R. POUNDER, *On the order of random channel networks*, SIAM J. Alg. Disc. Meth., 1 (1980), pp. 25–33.

[23] O. MESA, *Analysis of channel networks parameterized by elevation*, Ph.D. Dissertation, Dept. of Civil Engineering, University of Mississippi, University, MS, 1986.

[24] R. L. SHREVE, *Statistical law of stream numbers*, J. Geol., 74 (1966), pp. 17–37.

[25] D. A. STEINGRAEBER, L. J. KASCHT, AND D. H. FRANCK, *Variation of shoot morphology and bifurcation ratio in sugar maple (Acer Saccarum) saplings*, Amer. J. Botany, 66 (1979), pp. 441–445.

[26] M. VAUCHAUSSADE AND G. VIENNOT, *Enumeration of RNA's secondary structures by complexity*, in Proc. Inter. Conference Math. in Medicine and Biology, Bary, Italy, 1983.

[27] C. WERNER, *Two models for Horton's law of stream numbers*, Canadian Geographer, 16(1), pp. 50–68.