<table>
<thead>
<tr>
<th>Title</th>
<th>The number of endpoints in a random recursive tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Maehara, Hiroshi</td>
</tr>
<tr>
<td>Citation</td>
<td>琉球大学教育学部紀要 第二部 1984-01 109-112</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1984-01</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/20.500.12000/9631">http://hdl.handle.net/20.500.12000/9631</a></td>
</tr>
<tr>
<td>Rights</td>
<td></td>
</tr>
</tbody>
</table>
The number of endpoints in a random recursive tree

Hiroshi MAEHARA*

(Received August 20, 1983)

Abstract

It is proved that the distribution of the number of endpoints in a random recursive tree of order \( n \) is asymptotically normal \((n \to \infty)\) with asymptotic mean \( n/2 \) and asymptotic variance \( n/12 \).

1. Introduction

A tree is a connected graph that has no cycle. The order of a tree is the number of points in the tree. In a labeled tree of order \( n \), the integers 1 through \( n \) are assigned to its points. A labeled tree of order \( n \) is called a recursive tree (see Moon [3]) if \( n=1 \), or \( n \geq 2 \) and it is obtained by joining the \( n \)th point to one point of some recursive tree of order \( n-1 \). Thus, in a recursive tree the labeling is considered to show the process of growth. And it is easily seen that there are exactly \((n-1)!\) recursive trees of order \( n \).

A point of a tree is an endpoint if its degree is one. Let \( V_n \) be the number of endpoints in a random labeled tree of order \( n \), that is, in a tree chosen at random from the set of all labeled trees of order \( n \). Then it was proved by Rényi [4] that the distribution of \( V_n \) is asymptotically normal as \( n \to \infty \) with asymptotic mean \( n/e \) and asymptotic variance \((2-e)n/e^2\).

In this note we consider the corresponding problem for a random recursive tree. Let \( X(n) \) be the number of endpoints in a recursive tree of order \( n \) which is chosen at random from the set of all recursive trees of order \( n \). For a technical reason, however, we exclude the first point (= the point of label 1) in the count of endpoints, even if its degree is one. Thus \( X(1) = 0 \), \( X(2) = 1 \) (!), and \( X(3) \), \( X(4) \), \( \ldots \ldots \) are random variables. We shall show that the distribution of \( X(n) \) is asymptotically normal as \( n \) tends to infinity with asymptotic mean \( n/2 \) and asymptotic variance \( n/12 \).

2. The conditional probability

Let \( T(1), T(2), T(3), \ldots \) be the sequence of random recursive trees defined inductively in the following way:

(i) \( T(1) \) is the single point 1,

and

(ii) for \( n \geq 2 \), \( T(n) \) is a random recursive tree of order \( n \) obtained by joining the \( n \)th point to one point randomly selected from the \( n-1 \) points of \( T(n-1) \).

* College of Education, Ryukyu University.
Then it is easily proved by mathematical induction on \( n \) that each recursive tree of order \( n \) is achieved by \( T(n) \) with equal probability. Hence we regard \( X(n) \) as the number of endpoints of \( T(n) \), for each \( n \). (Of course the first point is excluded from the set of endpoints.) By doing so, \( X(n) \) and \( X(n+1) \) are related as follows. If \( T(n) \) has \( x \) endpoints then \( T(n+1) \) is to have \( x \) or \( x+1 \) endpoints accordingly as the point joined with the \((n+1)\)th point is or is not an endpoint of \( T(n) \). Thus the conditional probability of \( X(n+1) = y \) on the hypothesis \( X(n) = x \) is given by

\[
P(X(n+1) = y \mid X(n) = x) = \begin{cases} \frac{x}{n} & \text{if } y = x \\ \frac{1-x}{n} & \text{if } y = x+1 \\ 0 & \text{otherwise.} \end{cases}
\]

Indeed, \( \{X(n) ; n=1, 2, 3, \cdots \} \) is a nonstationary Markov process (see e.g. [1, p.369]) with one step transition probability (2.1) at 'time' \( n \).

3. The mean and variance

Let us denote by \( \mu(n) \) and \( \sigma(n)^2 \) the mean and variance of the random variable \( X(n) \). We shall show that

\[
\mu(n) = \frac{n}{2} \quad \text{for } n \geq 2,
\]

and

\[
\sigma(n)^2 = \frac{n}{12} \quad \text{for } n \geq 3.
\]

For random variables \( X \) and \( Y \), we denote by \( \mathbb{E} \{ X \mid Y = y \} \) the conditional expectation of \( X \) given \( Y = y \), and by \( \mathbb{E} \{ X \mid Y \} \) that function of the random variable \( Y \) whose value at \( Y = y \) is \( \mathbb{E} \{ X \mid Y = y \} \).

Now, from the conditional probability (2.1), we have

\[
\mathbb{E} \{ X(n+1) \mid X(n) = x \} = x \left( \frac{x}{n} \right) + (x+1) \left( 1 - \frac{x}{n} \right) = (1 - \frac{1}{n})x + 1.
\]

Hence

\[
\mathbb{E} \{ X(n+1) \} = \mathbb{E} \{ \mathbb{E} \{ X(n+1) \mid X(n) \} \} = (1 - \frac{1}{n}) \mathbb{E} \{ X(n) \} + 1.
\]

Since \( \mu(2) = \mathbb{E} \{ X(2) \} = 1 \), (3.1) follows easily from this recursion formula.

For an integer \( h \geq 0 \), let \( \mu_h(n) \) denote the \( h \)th central moment of \( X(n) \). If we put \( Y(n) = X(n) - \mu(n) \) then

\[
\mu_h(n+1) = \mathbb{E} \{ Y(n+1)^h \} = \mathbb{E} \{ \mathbb{E} \{ Y(n+1)^h \mid X(n) \} \}
\]

\[
= \mathbb{E} \{ (X(n) - \mu(n+1))^h \} X(n)/n + (X(n)+1-\mu(n+1))^h \{ 1-X(n)/n \}
\]

\[
= \mathbb{E} \{ (Y(n) - 1/2)^h \} \{ Y(n)/n+1/2 \} - \{ Y(n)+1/2 \}^h \{ Y(n)/n-1/2 \}.
\]

After some calculations we get the following recursion formula:

\[
\mu_h(n+1) = \sum_{i \text{ even}} 2^{-i} \left( \binom{h}{i} - \frac{1}{n} \binom{h}{i+1} \right) \mu_{h-i}(n),
\]

\[\text{-110-}\]
MAEHARA: The number of endpoints in a random recursive tree

where the sum is over all even numbers \( i \), \( 0 \leq i \leq h \).

Letting \( h = 2 \) yields

\[
\mu_2(n+1) = (1 - 2/n) \mu_2(n) + 1/4,
\]

from which we find that

\[
\sigma(n)^2 = \mu_2(n) = n/12 \quad \text{for } n \geq 3.
\]

4. The higher moments

We will show that for each fixed integer \( j \geq 0 \)

\[
(4.1) \quad \mu_{2j}(n) = \frac{(2j)!}{2^j j!} \left( \frac{n}{12} \right)^j + O(n^{j-1}) \quad (n \to \infty)
\]

and

\[
(4.2) \quad \mu_{2j+1}(n) = 0 \quad \text{for all } n \geq 2.
\]

First we show \((4.2)\) by induction on \( j \). Obviously \( \mu_1(n) = 0 \). Suppose \( \mu_{2j+1}(n) = 0 \) for \( j < k \), \( k \geq 1 \). Then from \((3.3)\) we have

\[
\mu_{2k+1}(n+1) = (1 - (2k+1)/n) \mu_{2k+1}(n).
\]

Since \( \mu_{2k+1}(2) = 0 \), it follows that \( \mu_{2k+1}(n) = 0 \) for \( n \geq 2 \).

To prove \((4.1)\) we prepare a lemma.

**Lemma.** Let \( \ell \), \( m \geq 0 \) be fixed integers and \( g(n) \) be a function on \( n \) such that \( g(n) = cn^t + O(n^{t-1}) \), where \( c \) is a constant. If \( f(n) \) satisfies the recurrence relation

\[
f(n+1) = (1 - m/n) f(n) + g(n)
\]

then

\[
f(n) = \frac{c}{\ell + m + 1} n^{t+1} + O(n^t).
\]

**Proof.** From the recurrence relation, we successively see that

\[
f(m+1) = g(m),
\]

\[
f(m+2) = \frac{1}{m+1} g(m) + g(m+1),
\]

\[
f(m+3) = \frac{2 \cdot 1}{(m+2)(m+1)} g(m) + \frac{2}{m+2} g(m+1) + g(m+2),
\]

\[\ldots\ldots\]

Thus we find that

\[
f(n+1) = \sum_{i=0}^{n-m} \frac{(m+i)! (n-m)!}{n! i!} g(m+i) = \frac{1}{(n)_m} \sum_{i=0}^{n} (i)_m g(i)
\]

where \((x)_m\) denotes the falling \( m \)-factorial of \( x \), that is,

\[(x)_m = x(x-1) \cdots (x-m+1).\]
Then the lemma follows from
\[(x)_m = x^m + O(x^{m-1})\]
and
\[\sum_{i=m}^{n} i^{t+m} = \frac{1}{t+m+1} n^{t+m+1} + O(n^{t+m}).\]

Now we show (4.1) by induction on \(j\). Since \(\mu_k(n) = 1\), (4.1) holds for \(j = 0\). Suppose (4.1) holds for \(j < k\), \(k \geq 1\). Then by (3.3) we have
\[\mu_{2k}(n+1) = (1 - 2k/n)\mu_{2k}(n) + \frac{1}{4} \binom{2k}{2} - \frac{1}{4n} \binom{2k}{3} \mu_{2k-2}(n) + \cdots \]
\[= (1 - 2k/n)\mu_{2k}(n) + \frac{1}{4} \binom{2k}{2} \frac{(2k-2)!}{2^{k-1}(k-1)!} \left(\frac{n}{12}\right)^{k-1} + O(n^{k-2}).\]
Applying the lemma we have
\[\mu_{2k}(n) = \frac{1}{k-1 + 2k + 1} \cdot \frac{1}{4} \binom{2k}{2} \frac{(2k-1)!}{2^{k-1}(k-1)!} \left(\frac{n}{12}\right)^{k-1} n + O(n^{k-1})\]
\[= \frac{(2k)!}{2^k k!} \left(\frac{n}{12}\right)^k + O(n^{k-1}).\]
Thus (4.1) also holds for \(j = k\).

5. The asymptotic distribution

We now prove the distribution of \((X(n) - n/2)/\sigma(n)\) tends to the normal distribution with zero mean and unit variance. To do this it is sufficient [2, p.115] to show that for each fixed integer \(j\),
\[\frac{\mu_{2j}(n)}{\sigma(n)^{2j}} \to \frac{(2j)!}{2^j j!} \quad \text{and} \quad \frac{\mu_{2j+1}(n)}{\sigma(n)^{2j+1}} \to 0\]
as \(n\) tends to infinity. But this follows at once from (3.2) and (4.1), (4.2).

Thus the distribution of \(X(n)\) is asymptotically normal \((n \to \infty)\) with asymptotic mean \(n/2\) and asymptotic variance \(n/12\).

References