arXiv:2305.03690v1 [math.PR] 5 May 2023

# 按叶子数计算分枝过程树的子树数量

BORIS PITTEL

Abstract. 我们研究在灭绝时，从临界分枝过程树的所有子树中均匀随机选取一棵子树时，其叶子数的分布。

## 1. Introduction and results

考虑一个由单个始祖启动的分枝过程。这个过程可以形象地表示为一棵不断生长的有根树。根节点是始祖，并通过边与它的每个直接后代（子节点）相连，这些子节点按某种顺序排列，例如按长幼顺序。每个子节点都成为相应子树的根节点，因此所有这些根节点的子节点便是始祖的孙辈。显然，我们得到的是一个递归定义的过程。它产生一列嵌套的树，这一序列要么无限延续，要么在某一时刻终止，此时当前所有叶子都没有子节点。

经典的 Galton-Watson 分枝过程是指：每个成员的子节点数目 (a) 独立于前一代和当前一代所有成员的这些数目，并且 (b) 具有相同的分布 $\{p_j\}_{j\ge 0}$，($\sum_j p_j = 1$)。众所周知，如果 $p_0 > 0$ 和$\sum_{j\ge 0} jp_j = 1$，则该过程以概率 1 终止，Harris [7]。一个标准例子是 $p_0 = p_2 = \frac{1}{2}$，在这种情况下我们得到一棵二叉树。

令 $T$ 表示终端树。给定一棵有限根树 $\mathcal{T}$，我们有

$$\mathbb{P}(T = \mathcal{T}) = \prod_{v\in\mathcal{V}(\mathcal{T})} p_{d(v,\mathcal{T})},$$

其中，$\mathcal{V}(\mathcal{T})$ 是 $\mathcal{T}$ 的顶点集，$d(v,\mathcal{T})$ 是顶点 $v \in V(\mathcal{T})$ 的出度。定义 $L = L_T$ $T$ 的叶子总数，即 $T$ 中出度为 0 的顶点数；并令 $X(t) = X_T(t)$ 表示 $T$ 中具有 $t$ 个叶子的（满）子树总数。因此，子树总数 $\sum_{t\ge 1} X(t)$ $= V(T)$ $=|\mathcal{V}(T)|$。定义 $\mathcal{L} = \mathcal{L}(T)$ 为从 $T$ 的所有 $V = V(T)$ 子树中均匀随机选取一个子树时，其中叶子的数量；因此，在给定 $T$ 的条件下，有 $\mathbb{P}$ $(\mathcal{L} = t|T)$ $= {}^{X}(t)\,{}_{V}$；而仅在给定 $L$ 的条件下，$\mathbb{P}(\mathcal{L} = t|L)$ $=$

2020 *Mathematics Subject Classification.* 60C05; 05C05, 92B10.

*Key words and phrases.* 分枝过程、随机树、灭绝、渐近性。1

$\mathbb{E}\big[\frac{X(t)}{V}|L\big]$。 [15] 中已证明， $f(x) := \mathbb{E}[x^L] = \sum_{k\ge1} x^k\mathbb{P}(L = k)$ 满足

$$f(x) = \sum_{j\ge1} p_j f^j(x) + p_0 x, \quad |x| \le 1. \tag{1.1}$$

此外， $\mathbb{P}(L < \infty) = 1$，并假设 $\{p_j\}$ 具有有限方差 $\sigma^2$,

$$\mathbb{P}(L = \ell) = \frac{\gamma(2\ell-1)!!}{2^\ell \ell!} + O(\ell^{-2}) = \frac{\gamma}{\ell^{3/2}} + O(\ell^{-2}), \quad \gamma := \Big(\frac{p_0}{2\pi\sigma^2}\Big)^{1/2}, \tag{1.2}$$

$\sigma^2$ 为 $\{p_j\}$ 的方差。（需要说明的是，我们关注的是 $p_1 = 0$ 的情况，但公式 (1.2) 对所有 $p_1 < 1$ 同样成立。）一个精确性稍低的公式 $\mathbb{P}(L =\ell) = \frac{\gamma}{\ell^{3/2}}(1 + o(1))$ 早已为人所知，见 Kolchin（第 2 章，引理 4）。

定理 1.1. (i) 对于二元情形 $p_0 = p_2 = 1/2$,

$$\mathbb{P}(\mathcal{L} = t|L = \ell) = \frac{t}{\ell(2\ell-1)} \cdot \frac{\binom{\ell}{t}^2}{\binom{2(\ell-1)}{2(t-1)}}.$$

(ii) 因此，对于 $t = o(\ell)$，我们有

$$\mathbb{P}(\mathcal{L} = t|L = \ell) = \frac{1+O(\ell^{-1/2}+t/\ell)}{2^{2t-1}t}\binom{2(t-1)}{t-1} = (1 + O(\ell^{-1/2} + t/\ell))\mathbb{P}(L = t),$$

这意味着分布序列 $\{\mathbb{P}(\mathcal{L} = t|L = \ell)\}_{t\ge1}$ 的紧性。然而，
$\mathbb{E}[\mathcal{L}|L = \ell] \sim \frac{\sqrt{\pi\ell}}{2}$，因为 $\ell \to \infty$。

注。因此，对于较大的 $\ell$，从具有 $\ell$ 个叶子的灭绝树的所有 $(2\ell - 1)$ 个子树中随机选取的子树，其叶子数渐近地服从灭绝树的叶子数分布。

我们的第二个结果适用于一般的临界分布 $\{p_j\}_{j\ge0}$。

定理 1.2. (i) 对每个固定的 $t \ge 1$,

$$\lim_{\ell\to\infty} \mathbb{P}(\mathcal{L} = t|L = \ell) = (1 - p_1)\mathbb{P}(L = t).$$

由于 $\sum_{t\ge1} \mathbb{P}(L = t) = 1$，我们看到分布$\{\mathbb{P}(\mathcal{L} = t|L = \ell)\}_{1\le t\le\ell}$、($\ell \ge 1$)这一序列并不紧，除非（如同在二元情形中）$p_1 = 0$。换言之， $p_1$ 是随机子树的叶集大小分布的极限亏损。(ii) 更准确地说，

$$\mathbb{P}\Big(\mathcal{L} > \frac{\ell^{1/2}}{\log^2\ell}\Big|L = \ell\Big) = p_1 + O(\ell^{-1/4}\log\ell),$$

因此，在给定 $\{L = \ell\}$ 的条件下， $\mathcal{L}$ 超过 $\frac{\ell^{1/2}}{\text{对数}^2\,\ell}$ 的条件概率有一个与零保持正距离的下界，当 $\ell \to \infty$ 时。

注：由于 $p_1 < 1$，方程（1.1）等价于

$$f(x) = \sum_{j\ge2} p'_j f^j(x) + p'_0 x, \quad p'_j = \frac{p_j}{1-p_1},\ j = 0, 2, 3, \dots,$$

且 $p'_0$、$p'_2$、$p'_3$、.....构成一个概率分布，其中 $\sum_j jp'_j = 1$ 同样成立，但此时 $p'_1 = 0$。因此， $L$ 和 $L'$（即与 $\{p'_j\}$ 相关联的终端树 $T'$ 中叶子的数量）是同分布的。因此，(i) 部分可以解释为：在给定事件“没有任何父节点只有一个子节点”发生的条件下， $T$ 的均匀随机子树在极限情况下，其叶子数量的分布与树 $T'$ 的叶子数量分布相同。

作为我们的灵感来源，我们应当提及俄罗斯数学家 Valentin Kolchin。自20世纪70年代中期以来，他率先开展并积极推动了条件分枝过程与随机树组合学之间联系的研究，参见 Kolchin [9], [10], 和 [11]。关于条件树和简单生成树的极限结果（未涉及收敛速率），我们请读者参阅 David Aldous [1] 和 Svante Janson [8] 的两篇相隔20年的百科全书式综述，其中分析了丰富多样的边缘分布。随机子树大小的分布在 [1] 中被列为此类分布中的基本问题之一。

## 2. Generating functions identities

以下是对 $f(x)$ 应用 (1.1) 的隐式微分后得到的两个有用恒等式：

$$\sum_{j\geq 1} jp_j f^{j-1}(x) = 1 - \tfrac{p_0}{f'(x)}, \quad \sum_{j\geq 1} j(j-1)p_j f^{j-2}(x) = \tfrac{p_0 f''(x)}{(f'(x)^3}. \tag{2.1}$$

特别地，由于 $f(0) = 0$，我们有 $f'(0) = \frac{p_0}{1-p_1}$。

我们的首要任务是推导一个关于 $f(\ell, t) = \mathbb{E}[X(t)\mathbb{I}(L = \ell)]$, 的方程，其中 $\mathbb{I}(B)$ 表示事件 $B$ 的示性函数。当然，对于 $\ell < t$， $f(\ell, t) = 0$，且 $f(t, t) = \mathbb{P}(L = t)$。考虑 $\ell > t$。以概率 $p_j$，根节点有 $j$ 个子节点；令 $X_i(t)$ 表示以第 $i$ 个子节点为根的树中含有 $t$ 个叶子的子树数量。那么

$$\begin{aligned} f(\ell,t) = \mathbb{E}\big[X(t)\mathbb{I}(L=\ell)\big] &= \sum_{j\geq 1} p_j \mathbb{E}\Big[\Big(\sum_{i\in[j]} X_i(t)\Big)\cdot \mathbb{I}\Big(\sum_{i'\in[j]} L_{i'} = \ell\Big)\Big] \\ &= \sum_{j\geq 1} jp_j \mathbb{E}\Big[X_1(t)\cdot \mathbb{I}\Big(\sum_{i\in[j]} L_i = \ell\Big)\Big] \\ &= \sum_{j\geq 1} jp_j \sum_{\mu\leq \ell} \mathbb{E}\big[X(t)\mathbb{I}(L=\mu)\big]\cdot \mathbb{P}\Big(\sum_{i\in[j]\setminus\{1\}} L_i = \ell-\mu\Big) \\ &= \sum_{j\geq 1} jp_j \sum_{\mu\leq \ell} f(\mu,t)\cdot [x^{\ell-\mu}] f^{j-1}(x) \\ &= \sum_{\mu\leq \ell} f(\mu,t)\big[x^{\ell-\mu}\big] \sum_{j\geq 1} jp_j f^{j-1}(x). \end{aligned}$$

因此，结合（2.1），我们得到

$$f(\ell,t)=\sum_{\mu\le\ell} f(\mu,t)\cdot\big[x^{\ell-\mu}\big]\big(1-\tfrac{p_0}{f'(x)}\big),\quad \ell>t. \tag{2.2}$$

并且我们提醒，$f(t,t)=\mathbb{P}(L=t)$。式 (2.2) 右端的卷积明确要求使用生成函数。对于 $|y|<1$，利用$f'(0)=\frac{p_0}{1-p_1}$，我们有

$$\begin{aligned}\sum_{\ell\ge t} y^\ell f(\ell,t) &= y^t\,\mathbb{P}(L=t)-y^t\,\mathbb{P}(L=t)\cdot[x^0]\big(1-\tfrac{p_0}{f'(x)}\big)\\ &\qquad+\sum_{\ell\ge t} y^\ell\sum_{\mu=t}^{\ell} f(\mu,t)\cdot[x^{\ell-\mu}]\big(1-\tfrac{p_0}{f'(x)}\big)\\ &=(1-p_1)y^t\,\mathbb{P}(L=t)+\sum_{\mu\ge t} f(\mu,t)y^\mu\sum_{\ell\ge\mu} y^{\ell-\mu}\cdot[x^{\ell-\mu}]\big(1-\tfrac{p_0}{f'(x)}\big)\\ &=(1-p_1)y^t\,\mathbb{P}(L=t)+\big(1-\tfrac{p_0}{f'(y)}\big)\sum_{\mu\ge t} f(\mu,t)y^\mu,\end{aligned}$$

这意味着

$$\sum_{\ell\ge t} y^\ell f(\ell,t)=\tfrac{1-p_1}{p_0}y^t f'(y)\,\mathbb{P}(L=t). \tag{2.3}$$

回想起 $f(\ell,t)=f(\ell,t)=\mathbb{E}[X(t)\mathbb{I}(L=\ell)]$,并使用 $[y^s]f'(y)=(s+1)\cdot[y^{s+1}]f(y)$，我们得到

引理 2.1.

$$\mathbb{E}[X(t)|L=\ell]=\tfrac{1-p_1}{p_0}\tfrac{[y^{\ell-t}]f'(y)\times\mathbb{P}(L=t)}{\mathbb{P}(L=\ell)}=\tfrac{1-p_1}{p_0}\,\tfrac{(\ell-t+1)\mathbb{P}(L=\ell-t+1)\,\mathbb{P}(L=t)}{\mathbb{P}(L=\ell)}.$$

## 3. Proof of Theorem 1.1

注意，在二元情形下，事件 $\{L=\ell\}$ 上有 $V=2\ell-1$。

证明。根据引理 2.1（第一个恒等式），经过一些初等运算，我们得到

$$\begin{aligned}\mathbb{P}(\mathcal{L}=t|L=\ell)&=\tfrac{1}{2\ell-1}\mathbb{E}[X(t)|L=\ell]=\tfrac{2}{2\ell-1}\tfrac{[y^{\ell-t}]f'(y)\times[x^t]f(x)}{[x^\ell]f(x)}\\ &=\tfrac{1}{2\ell-1}\cdot\tfrac{[y^{\ell-t}](1-y)^{-1/2}\times[x^t](-(1-x)^{1/2})}{[x^\ell](-(1-x)^{1/2})}=\tfrac{t}{\ell(2\ell-1)}\cdot\tfrac{\binom{\ell}{t}^2}{\binom{2(\ell-1)}{2(t-1)}}.\end{aligned}$$

渐近公式 $\mathbb{E}[\mathcal{L}|L=\ell]\sim\frac{\sqrt{\pi\ell}}{2}$ 可由此直接得到。

$$t\cdot\mathbb{P}(\mathcal{L}=t|L=\ell)=\tfrac{\ell}{2\ell-1}\tfrac{\binom{2(t-1)}{t-1}\binom{2(\ell-t)}{\ell-t}}{\binom{2(\ell-1)}{\ell-1}},$$

公式 $\binom{2m}{m} \sim \frac{2^{2m}}{\sqrt{\pi m}}$、$(m \to \infty)$ 和 $\int_0^1 \frac{dx}{\sqrt{x(1-x)}} = \pi$。最后，根据 (1.2) 和引理 2.1 中的第二个恒等式，我们有

$$\begin{aligned}\mathbb{P}(\mathcal{L} = t|L = \ell) &= \tfrac{1}{2\ell-1}\mathbb{E}[X(t)|L = \ell] \\ &= \tfrac{2(\ell-t+1)(\ell-t+1)^{-3/2}}{(2\ell-1)\ell^{-3/2}} \cdot \big(1 + O((\ell - t + 1)^{-1/2})\big)\mathbb{P}(L = t) \\ &= (1 + O(\ell^{-1/2} + t/\ell))\mathbb{P}(L = t),\end{aligned}$$

前提是 $t = o(\ell)$。 □

## 4. Proof of Theorem 1.2

对于一般的 $\{p_j\}_{j\geq 0}$，给定 $p_0 > 0$、$\sum_j jp_j = 1$、$V = \sum_s X(s)$ 在事件 $\{L = \ell\}$ 上是随机的，即可供选择的灭绝树的子树数量是随机的。因此，这里 $\mathbb{P}(\mathcal{L} = t|L = \ell) = \mathbb{E}\big[\frac{X(t)}{V}\big|L = \ell\big]$. 。对依赖随机变量的比值进行精确甚至渐近的矩评估可能会遇到问题。幸运的是，我们希望，对于 $\ell$ 较大时，在给定 $\{L = \ell\}$ 的条件下，$V$ 的分布会高度集中在 $\mathbb{E}[V|L = \ell]$ 附近。如果确实如此，那么我们可以预期，对于较大的 $\ell$，$\mathbb{P}(\mathcal{L} = t|L = \ell) \sim \frac{\mathbb{E}[X(t)|L=\ell]}{\mathbb{E}[V|L=\ell]}$ 对于相当广泛的 $t < \ell$ 都成立。而最后一个分式非常适合进行渐近分析。

为证明集中性，我们引入 $g(y, \ell) = \mathbb{E}[y^V \mathbb{I}(L = \ell)], (\ell \geq 1)$，使得 $g(y, 1) = p_0 y$，并定义 $G(y, x) := \sum_{\ell\geq 1} x^\ell g(y, \ell)$。对于 $\ell > 1$，我们有

$$g(y, \ell) = y \sum_{j\geq 1} p_j \sum_{\ell_1+\cdots+\ell_j=\ell} \prod_{i\in[j]} g(y, \ell_i) = [x^\ell]\, y \sum_{j\geq 1} p_j G^j(y, x).$$

因此

$$G(y, x) = p_0 yx + y \sum_{j\geq 1} p_j G^j(y, x). \tag{4.1}$$

我们使用 (4.1) 得到 $G'_y(1, x)$ 和 $G''_y(1, x)$，因为

$$\begin{aligned} &G'_y(1, x) = \mathbb{E}[Vx^L] \Longrightarrow \mathbb{E}[V\mathbb{I}(L = \ell)] = [x^\ell]G'_y(1, x), \\ &G''_y(1, x) = \mathbb{E}[V(V - 1)x^L] \Longrightarrow \mathbb{E}[V(V - 1)\mathbb{I}(L = \ell)] = [x^\ell]G''_y(1, x).\end{aligned} \tag{4.2}$$

我们有

$$G(1,x)=\sum_{\ell\ge1}x^\ell g(1,\ell)=\sum_{\ell\ge1}x^\ell\mathbb{P}(L=\ell)=f(x),$$

$$\begin{aligned}G'_y(1,x)&=\Big(p_0x+\sum_{j\ge1}p_jG^j(y,x)+y\sum_{j\ge1}jp_jG^{j-1}(y,x)G'_y(y,x)\Big)\Big|_{y=1}\\&=f(x)+\Big(\sum_{j\ge1}jp_jf^{j-1}(x)\Big)G'_y(1,x),\end{aligned}$$

$$\begin{aligned}G''_y(1,x)&=\Big(2\sum_{j\ge1}jp_jG^{j-1}(y,x)G'_y(y,x)\\&\quad+y\sum_{j\ge1}jp_j\big[(j-1)G^{j-2}(y,x)(G'_y(y,x))^2+G^{j-1}(y,x)G''_y(y,x)\big]\Big)\Big|_{y=1}\\&=2G'_y(1,x)\sum_{j\ge1}jp_jf^{j-1}(x)\\&\quad+(G'_y(1,x))^2\sum_{j\ge1}j(j-1)p_jf^{j-2}(x)+G''_y(1,x)\sum_{j\ge1}jp_jf^{j-1}(x).\end{aligned}$$

因此

$$\begin{aligned}G'_y(1,x)&=\tfrac{f(x)}{1-\sum_j jp_jf^{j-1}(x)},\\G''_y(1,x)&=\frac{\sum_j j(j-1)p_jf^j(x)}{\Big(1-\sum_j jp_jf^{j-1}(x)\Big)^3}+\frac{2\Big(\sum_j jp_jf^j(x)\Big)}{\Big(1-\sum_j jp_jf^{j-1}(x)\Big)^2}.\end{aligned}\tag{4.3}$$

因此，结合（4.2）、（4.3）和（2.1），我们得出结论：

$$\begin{aligned}\mathbb{E}[V\mathbb{I}(L=\ell)]&=[x^\ell]\tfrac{f(x)f'(x)}{p_0},\\\mathbb{E}[V(V-1)\mathbb{I}(L=\ell)]&=\tfrac{1}{p_0^2}\cdot[x^\ell]\big(f''(x)f^2(x)\big),\\&\quad+2\cdot[x^\ell]\big(\tfrac{(f'(x))^2f(x)}{p_0^2}-\tfrac{f'(x)f(x)}{p_0}\big).\end{aligned}\tag{4.4}$$

让我们计算这些期望。现在来做一些复分析。利用魏尔斯特拉斯预备定理（参见 Ebeling [4], Krantz 和 Parks [12]），我们证明了 In [15] that $f(z)=\sum_{k\ge1}z^k\,\mathbb{P}(L=k)$, $z\in\mathbb{C}$, $|z|<1$ 在以原点为中心、半径为$\rho>1$的开圆盘 $D$（去掉一条割线 $[1,\rho)$）上存在解析延拓 $F(z)$，使得对于 $z\in D\setminus[1,\rho)$, $z\to1$,

$$F(z)=1-\gamma(1-z)^{1/2}+\sum_{s>1}\gamma_s(1-z)^{s/2},\ \gamma:=(2p_0)^{1/2}\sigma^{-1}.\tag{4.5}$$

这里 $\xi^{1/2}:=|\xi|^{1/2}\exp(i\mathrm{Arg}(\xi)/2)$, $\mathrm{Arg}(\xi)\in(-\pi,\pi)$。因此，在略微高于割线 $\{z=x,x\in[1,\rho)\}$ 的位置，我们有 $(1-z)^{1/2}=|1-z|^{1/2}e^{-i\pi/2}=$

$-i|1-x|^{1/2}$。（我们鼓励感兴趣的读者查阅 Bender [2], Can- field [3], Meir 和 Moon [13], [14] 以及 Flajolet 和 Sedgewick [5], ，了解经典复分析技术如何融入解析组合学这一精彩历程。）

由于 $\frac{F}{(z)F'(z)\,z^{\ell+1}}$ 在圆形围道 $\mathcal{C}_r = \{z = re^{i\theta},\, \theta \in (-\pi, \pi]\}$, $r = 1$ 上可积，我们利用柯西积分定理写出

$$[x^\ell]\big(f(x)f'(x)\big) = \tfrac{1}{2\pi i}\oint_{\mathcal{C}_1} \tfrac{F(z)F'(z)}{z^{\ell+1}}\,dz$$
$$= \tfrac{1}{2\pi i}\oint_{\mathcal{C}_\rho} \tfrac{F(z)F'(z)}{z^{\ell+1}} + \tfrac{1}{2\pi i}\int_1^\rho \tfrac{2i\,\mathrm{Im}(F(x)F'(x))}{x^{\ell+1}}\,dx.$$

对于第二行，我们用极限围道替换了 $\mathcal{C}_1$。它由定向圆 $z = \rho e^{i\alpha}$、$0 \le a \le 2\pi$ （其中挖去了单点 $z = \rho$ ）以及一个绕行部分组成，后者由两条方向相反的线段构成，一条从 $z = \rho e^{i(2\pi-0)}$ 到 $z = e^{i(2\pi-0)}$，另一条从 $z^{i(+0)}$ 到 $z = \rho^{i(+0)}$。出现 $2i\,\mathrm{Im}(F(x)F'(x))$ 是因为 $F(x)F'(x)$ 在 $[1, \rho)$ 的正上方和正下方的取值互为复共轭，因此它们的实部相互抵消，而 $\mathrm{Im}(F(x)F'(x))$ 则是由于 $z$ 从上方趋近于 $x \in [1, \rho)$ 而产生的。底部的第一个积分是 $O(\rho^{-\ell})$ 阶的。此外，利用 (4.5) 以及 Gradschteyn 和 Ryzik 中的公式 3.191(2) [6], ，我们得到

$$\tfrac{1}{2\pi i}\int_1^\rho \tfrac{2i\,\mathrm{Im}(F(x)F'(x))}{x^{\ell+1}}\,dx = \tfrac{\gamma}{2\pi}\int_1^\rho \tfrac{(x-1)^{-1/2}+O(1)}{x^{\ell+1}}\,dx$$
$$= \tfrac{\gamma}{2\pi}\int_1^\infty \tfrac{(x-1)^{-1/2}}{x^{\ell+1}}\,dx + O(1/\ell) = \tfrac{\gamma(2\ell-1)!!}{2^{\ell+1}\ell!} + O(1/\ell);$$

因此，显式项的阶恰好为 $\ell^{-1/2}$ 。

$$\mathbb{E}[V\mathbb{I}(L=\ell)] = \tfrac{\gamma(2\ell-1)!!}{p_0 2^{\ell+1}\ell!} + O(1/\ell). \tag{4.6}$$

现在，根据（1.2），我们有 $\mathbb{P}(L = \ell) = \frac{\gamma}{(2\ell-3)!!\ 2^\ell \ell!} + O(\ell^{-2})$。由此及（4.6）可知，

$$\mathbb{E}[V|L=\ell] = \tfrac{\mathbb{E}[V\mathbb{I}(L=\ell)]}{\mathbb{P}(L=\ell)} = \tfrac{\ell}{p_0} + O(\ell^{1/2}). \tag{4.7}$$

转而考察（4.4）中的第二个恒等式。与上述计算类似，我们得到

$$\begin{aligned}
&\tfrac{1}{p_0^2}\cdot[x^\ell]\big(f''(x)f^2(x)\big)=\tfrac{\gamma(2\ell+1)!!}{(2p_0)^2 2^\ell \ell!}+O(1),\\
&2\cdot[x^\ell]\big(\tfrac{(f'(x))^2 f(x)}{p_0^2}-\tfrac{f'(x)f(x)}{p_0}\big)=O(1),\\
\Longrightarrow&\mathbb{E}[V(V-1)\mathbb{I}(L=\ell)]=\tfrac{\gamma(2\ell+1)!!}{(2p_0)^2 2^\ell \ell!}+O(1),\\
\Longrightarrow&\mathbb{E}[V^2\mathbb{I}(L=\ell)]=\tfrac{\gamma(2\ell+1)!!}{(2p_0)^2 2^\ell \ell!}+O(1),\\
\Longrightarrow&\mathbb{E}[V^2|L=\ell]=\big(\tfrac{\ell}{p_0}\big)^2+O(\ell^{3/2}).
\end{aligned}\tag{4.8}$$

将 (4.7) 和 (4.8) 代入，得到

$$\begin{aligned}
\mathrm{var}(V|L=\ell)&:=\mathbb{E}\big[(V-\mathbb{E}[V|L=\ell])^2|L=\ell\big]\\
&=\mathbb{E}[V^2|L=\ell]-\mathbb{E}^2[V|L=\ell]=O(\ell^{3/2}).
\end{aligned}\tag{4.9}$$

回想一下，$V$ 是终端树 $T$ 中顶点的数量，而 $X(t)$ 是其后代树有 $t$ 片叶子的顶点数量。因此，$\{\frac{X(t)}{V}\}_{t\ge 1}$ 是 $Y$ 的分布，即以 $T$ 为条件时，均匀随机选取的顶点所根植的总后代子树中的叶子数量。我们利用 (4.9) 和 $X(t)\le \frac{\ell}{t}$ 来估计；这是因为任意两个各有 $t$ 片叶子的子树都不相交。

$$\mathbb{E}\bigg[\Big(\frac{X(t)}{V}-\frac{X(t)}{\mathbb{E}[V|L=\ell]}\Big)^2\,\Big|L=\ell\bigg]\le\mathbb{E}\Big[\big(\tfrac{X(t)}{\ell^2}\big)^2\mathrm{var}(V|L=\ell)\Big]=O(\ell^{-1/2}t^{-2}).$$

因此，根据柯西–施瓦茨不等式，对所有 $t\ge 1$ 一致地有：

$$\mathbb{E}\big[\tfrac{X(t)}{V}|L=\ell\big]-\tfrac{\mathbb{E}[X(t)|L=\ell]}{E[V|L=\ell]}=O(\ell^{-1/4}t^{-1}).$$

或者等价地

$$\mathbb{P}(\mathcal{L}=t|L=\ell)=\tfrac{\mathbb{E}[X(t)|L=\ell]}{E[V|L=\ell]}+O(\ell^{-1/4}t^{-1}).$$

将此公式与引理 2.1 和 (4.7) 结合，得到

$$\mathbb{P}(\mathcal{L}=t|L=\ell)=(1-p_1)\tfrac{(\ell-t+1)\,\mathbb{P}(L=\ell-t+1)\,\mathbb{P}(L=t)}{\ell\,\mathbb{P}(L=\ell)}+O(\ell^{-1/4}t^{-1}),\tag{4.10}$$

对 $t\le\ell$ 一致成立。将公式 (1.2) 应用于 (4.10)，我们得出

$$\lim_{\ell\to\infty}\mathbb{P}(\mathcal{L}=t|L=\ell)=(1-p_1)\mathbb{P}(A_t).$$

由于 $\sum_{t\ge1}\mathbb{P}(A_t)=1$，我们可以看到，分布序列 $\{\mathbb{P}(Y=t|L=\ell)\}_{1\le t\le\ell}$、($\ell\ge1$) 并不紧，除非（如同二元情形）$p_1=0$。换言之，$p_1$ 是随机子树的叶集大小分布的极限亏损。

让我们将这一论断表述得更精确些。引入 $\tau = \frac{\ell^{1/2}}{\log^2 \ell}$。对于 $t \le \tau$，由 (1.2) 可得

$$\frac{(\ell-t+1)\,\mathbb{P}(L=\ell-t+1)}{\ell\,\mathbb{P}(L=\ell)} = 1 + O(\tau\ell^{-1} + \ell^{-1/2}) = 1 + O(\ell^{-1/2}),$$

这意味着

$$\begin{aligned}(1-p_1)\sum_{t\le\tau}\frac{(\ell-t+1)\,\mathbb{P}(L=\ell-t+1)\,\mathbb{P}(L=t)}{\ell\,\mathbb{P}(L=\ell)} &= (1-p_1)\sum_{t\le\tau}\mathbb{P}(L=t) + O(\ell^{-1/2})\\ &= 1-p_1+O(\mathbb{P}(L>\tau)) + O(\ell^{-1/2}) = 1-p_1+O\big(\ell^{-1/4}\log\ell\big).\end{aligned}$$

因此

$$\begin{aligned}\mathbb{P}(\mathcal{L}>\tau|L=\ell) &= 1-(1-p_1)\sum_{t\le\tau}\frac{(\ell-t+1)\,\mathbb{P}(L=\ell-t+1)\,\mathbb{P}(L=t)}{\ell\,\mathbb{P}(L=\ell)}\\ &\quad + O(\ell^{-1/4}\log\ell) = p_1 + O(\ell^{-1/4}\log\ell).\end{aligned}$$

这就完成了定理 1.2 的证明。

致谢。我感谢 David Aldous 帮助我正确认识这项工作的意义。

## References

[1] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, 《应用概率年刊》1（1991）228–266。[2] E. A. Bender, *Asymptotic methods in enumeration*, 《SIAM评论》16（1974）485–515。[3] E. R. Canfield, *Remarks on an asymptotic method in combinatorics*, 《组合理论杂志 A辑》37（1984）348–352。[4] W. Ebeling, *Functions of Several Complex Variables and Their Singularities*, 美国数学学会（2007）。[5] P. Flajolet 和 R. Sedgewick, *Analytic Combinatorics*, 剑桥大学出版社（2009）。[6] L. S. Gradshteyn 和 I. M. Ryzhik, *Table of Integrals, Series, and Products, 6th Edition*, 学术出版社（2000）。[7] T. E. Harris, *The Theory of Branching Processes*, 施普林格出版社（1963）。[8] S. Janson, *Simply generated trees. conditioned Galton–Watson trees, random allocations and condensation*, 《概率调查》9（2012）103–252。[9] V. F. Kolchin, *Branching processes, random trees, and a generalized scheme of arrangements of particles*, 《数学札记》21（1977）386–394。[10] V. F. Kolchin, *Moment of degeneration of a branching process and height of a randomtree*, 《数学札记》24（1978），954–961。[11] V. F. Kolchin, *Random Mappings*, Optimization Software, 纽约, （俄文原版译本）（1988）。[12] S. G. Krantz 和 H. R. Parks, *The Implicit Function Theorem: History, Theory, andApplications*, Birkhäuser（2013）。[13] A. Meir 和 J. W. Moon, *On an asymptotic method in enumeration*, 《组合理论杂志 A辑》51（1989）77–89。

[14] A. Meir 和 J. W. Moon, *Erratum: "On an asymptotic method in enumeration"*, 《组合理论杂志 A辑》52（1989）163。[15] B. Pittel, *Expected number of induced subtrees shared by two independent copies of arandom tree*, 《SIAM离散数学杂志》37（2023），1–16。
Department of Mathematics, Ohio State University, 231 West 18-th Avenue,Columbus OH 43210-1175
*Email address*: pittel.1@osu.edu