Труды Петрозаводского государственного университета

Серия "Математика" **Выпуск 1, 1993 г.**

УДК 519.554
Павлов Ю.Л.

О СЛУЧАЙНЫХ ДЕРЕВЬЯХ

Получены предельные распределения ряда характеристик случайных корневых деревьев, производящая функция которых удовлетворяет некоторым ограничениям.

В книге В.Ф.Колчина [1] показано, что для изучения графов, являющихся деревьями с помеченными вершинами, можно применять методы теории ветвящихся процессов. Рассмотрим множество $T_n^{(1)}$ всех корневых деревьев, некорневые вершины которых занумерованы числами $1,2,\dots,n$, а корень имеет номер 0. Любая вершина дерева соединяется с корнем единственным путем. Назовем высотой вершины число ребер, образующих этот путь, а максимальную из высот вершин дерева назовем высотой дерева. Считая ребра направленными от корня, назовем кратностью вершины число выходящих из нее ребер. Введем на $T_n^{(1)}$ равномерное распределение вероятностей, тогда различные числовые характеристики дерева становятся случайными величинами. Обозначим $\mu_r(t,T_n^{(1)})$ число вершин высоты t, имеющих кратность r, $\mu_r(T_n^{(1)})$ - число вершин кратности r, $\eta(T_n^{(1)})$ - максимальную кратность вершины, $\tau(T_n^{(1)})$ - высоту дерева, $\mu(t,T_n^{(1)})$ - число вершин высоты t.

Вместе с $T_n^{(1)}$ рассмотрим начинающийся с одной частицы однородный ветвящийся процесс $G^{(1)}$ с дискретным временем, в котором число потомков каждой частицы имеет распределение Пуассона с произвольным параметром $\lambda>0$. Пусть $\mu_r(t,G^{(1)})$ означает число частиц в момент t, имеющих ровно r прямых потомков, а $\nu(G^{(1)})$ равно общему числу частиц за все время эволюции процесса.

Введем матрицы $\|\mu_r(t,T_n^{(1)})\|$, $\|\mu_r(t,G^{(1)})\|$, $t,r=0,1,\dots,n$, и матрицу $M=\|m_r(t)\|$ такого же размера, составленную из целых неотрицательных чисел. В [1] получены теоремы о предельном поведении $\mu_r(T_n^{(1)})$, $\eta(T_n^{(1)})$, $\tau(T_n^{(1)})$, $\mu(t,T_n^{(1)})$ при $n\to\infty$. Доказательство этих результатов опирается на то, что эти величины являются функциями от $\mu_r(t,T_n^{(1)})$ и на связь, которая устанавливается между $T_n^{(1)}$ и $G^{(1)}$ в следующей лемме.

Лемма 1. Справедливо равенство

$$P\{\|\mu_r(t,T_n^{(1)})\| = M\} = P\{\|\mu_r(t,G^{(1)})\| = M \mid \nu(G^{(1)}) = n+1\}.$$

Этот подход в [1] использован и для изучения случайных деревьев с ограничениями на кратности вершин. Пусть R – множество целых неотрицательных чисел, содержащее 0 и не совпадающее с $\{0,1\}$. Обозначим $T_{n,R}^{(1)}$ подмножество $T_n^{(1)}$ тех деревьев, кратности вершин которых принимают значения только из R. Рассмотрим также ветвящийся процесс $G_R^{(1)}$, у которого число потомков каждой частицы может иметь значение только из R и имеет распределение, получающееся из распределения Пуассона с параметром λ путем введения соответствующей нормировки. Обозначая характеристики $T_{n,R}^{(1)}$ и $G_R^{(1)}$ аналогично предыдущему, нетрудно получить следующее утверждение.

Лемма 2 [1].
Если $P\{\nu(G_R^{(1)})=n+1\} > 0$, то

$$P\{\|\mu_r(t,T_{n,R}^{(1)})\| = M\} = P\{\|\mu_r(t,G_R^{(1)})\| = M \mid \nu(G_R^{(1)}) = n+1\}.$$

В статье [2] изложенный выше подход был применен для изучения плоских корневых деревьев с висячим корнем. В таких деревьях имеется фактически "сдвоенный" корень, поскольку вместе с корнем выделяется единственная вершина, смежная с ним. Пусть $T_n^{(2)}$ – множество всех таких деревьев, у которых число "обыкновенных" некорневых вершин равно n, то есть общее число вершин равно $n+2$. Назовем высотой вершины число ребер, составляющих путь от этой вершины до "сдвоенного" корня, то есть до вершины, смежной с корнем. Остальные характеристики $T_n^{(2)}$ определяются аналогично $T_n^{(1)}$. Вместе с $T_n^{(2)}$ введем ветвящийся процесс Гальтона-Ватсона $G^{(2)}$, в котором число потомков каждой частицы имеет геометрическое распределение с параметром λ. Сохраняя смысл принятых обозначений, сформулируем доказанные в

[2] утверждения, подобные леммам 1 и 2.

Лемма 3. Справедливо равенство

$$P\{\|\mu_r(t,T_n^{(2)})\| = M\} = P\{\|\mu_r(t,G^{(2)})\| = M \mid \nu(G^{(2)}) = n+1\}.$$

Лемма 4. Если $P\{\nu(G_R^{(2)})=n+1\} > 0$, то

$$P\{\|\mu_r(t,T_{n,R}^{(2)})\| = M\} = P\{\|\mu_r(t,G_R^{(2)})\| = M \mid \nu(G_R^{(2)}) = n+1\}.$$

Заметим, что методы доказательства результатов о $T_n^{(1)}$ и $T_n^{(2)}$, использованные в [1] и [2], сходны, а предельные распределения характеристик этих деревьев в большинстве случаев отличаются лишь значениями параметров. Это обстоятельство делает целесообразной попытку получения общих результатов, справедливых для различных классов деревьев. Аналогичные соображения имеют место и для лесов, поскольку существует сходство между лесами, составленными из деревьев типа $T_n^{(1)}$ (см., например, [3]), и лесами, составленными из деревьев типа $T_n^{(2)}$ [4].

В статье [5] были получены некоторые результаты о корневых деревьях, производящая функция $f(z)$ которых удовлетворяет соотношению $f(z) = \Gamma(f(z))$, где Γ – некоторый оператор. В [5] рассматривались удовлетворяющие этому соотношению помеченные деревья ($T_n^{(1)}$), плоские деревья с висячим корнем ($T_n^{(2)}$), плоские бинарные деревья с висячим корнем ($T_{n,R}^{(2)}$) при $R=\{0,2\}$, а также деревья с непомеченными вершинами и рекурсивные деревья.

Для получения результатов о корневых деревьях, обобщающих полученные ранее в работах [1,2], рассмотрим процесс Гальтона-Ватсона, в котором число потомков ξ каждой частицы имеет следующее распределение:

$$p\{\xi=k\} = p_k, \quad k=0,1,2,\ldots, \tag{1}$$

а производящая функция имеет вид

$$F(z) = \sum_{k=0}^{\infty} p_k z^k. \tag{2}$$

Обозначим $\mu_r(t,G)$ число частиц в момент t, имеющих ровно r прямых потомков, $\nu(G)$ – общее число частиц, существовавших в G до его вырождения, $\tau(G)$ – момент вырождения, $\mu_r(G)$ – число частиц, имеющих ровно r прямых потомков.

Мы будем говорить, следуя [1], что процесс G обладает свойством A, если существует процесс G_θ, у которого число потомков ξ_θ каждой частицы имеет распределение

$$P\{\xi_\theta = k\} = \theta^k p_k \,/\, F(\theta), \quad k=0,1,2,\ldots, \tag{3}$$

где $\theta>0$ таково, что $F(\theta) = \theta F'(\theta)$ и в точке θ конечна левая производная $F''(\theta)$. Процесс G_θ в этом случае является критическим с конечной дисперсией числа потомков одной частицы, равной $B_\theta = \theta^2 F''(\theta)/F(\theta)$. Характеристики процесса G_θ будем обозначать аналогично соответствующим процессу G случайным величинам с указанием в скобках G_θ.

Рассмотрим множество T_n всех корневых деревьев некоторого класса с n некорневыми вершинами (об ограничениях на классы деревьев будет сказано ниже). Введем на T_n равномерное распределение вероятностей и обозначим, по аналогии с предыдущим, $\mu_r(t,T_n)$, $\mu_r(T_n)$, $\eta(T_n)$, $\tau(T_n)$, $\mu(t,T_n)$ случайные величины, равные соответственно числу вершин высоты t и кратности r, числу вершин кратности r, максимальной кратности вершин, высоте дерева, числу вершин в слое t.

Введем матрицы $\|\mu_r(t,T_n)\|$, $\|\mu_r(t,G)\|$, $t,r = 0,1,\ldots,n$, и матрицу $M = \|m_r(t)\|$ такого же размера, составленную из целых неотрицательных чисел. Предположим, что для рассматриваемых деревьев из множества T_n справедливо соотношение

$$P\{\|\mu_r(t,T_n)\| = M\} = P\{\|\mu_r(t,G)\| = M \mid \nu(G) = n+1\}. \tag{4}$$

Легко видеть, что различные классы корневых деревьев, рассмотренные в леммах 1-4, являются частными случаями T_n. Мы будем далее рассматривать такие классы корневых деревьев, для которых существуют ветвящиеся процессы, обладающие свойствами A и (4).

Рассмотрим $\mu_r(T_n)$. Пусть $\xi^{(r)}$ - случайная величина, имеющая распределение:

$$P\{\xi^{(r)}=k\} = P\{\xi=k \mid \xi\neq r\}, \quad k=0,1,\ldots, \tag{5}$$

и пусть d - максимальный шаг распределения (1), а d_r - максимальный шаг распределения (5). Обозначим также

$$\sigma_r^2 = p_r(1-p_r-(1-p_r)^2 p_r \,/\, B_\theta). \tag{6}$$

Теорема 1. Если $p_r > 0$, то при $s\to\infty$ равномерно относительно целых k, для которых $u = (kd_r/d - np_r)/(\sigma_r\sqrt{n})$ лежит в любом конечном интервале, $n = sd_r$,

$$P\{\mu_r(T_n) = kd_r/d\} = \frac{d_r}{d\sigma_r\sqrt{2\pi n}}\, e^{-u^2/2}(1+o(1)).$$

Доказательство. В силу леммы 2.2.3 книги [1]

$$P\{\|\mu_r(t,G)\| = M \mid \nu(G) = n+1\} =$$

$$= P\{\|\mu_r(t,G_\theta)\| = M \mid \nu(G_\theta) = n+1\}, \tag{7}$$

а для критического процесса G_θ мы можем применить теорему 2.3.1 [1], поэтому, учитывая, что $\mu_r(T_n)$ есть функция от $\mu_r(t,T_n)$, утверждение теоремы 1 следует из (4).

Из теоремы 1 легко получить результаты о предельном поведении $\mu_r(T_n)$ для $T_n=T_n^{(1)}$ и $T_n=T_n^{(2)}$.

Заметим, что для критических ветвящихся процессов $d=d_r=1$, $B_\theta=1$ для распределения Пуассона и $B_\theta=2$ для геометрического. Отсюда следует теорема 2.5.1 [1] для $\mu_r(T_n^{(1)})$, а для $\mu_r(T_n^{(2)})$ получаем следующий результат.

Следствие 1. Пусть $p_r=2^{-(r+1)}$, $B_\theta=2$. Если $n\to\infty$, то для любого фиксированного r равномерно относительно целых $k=np_r+u\sigma_r\sqrt{n}$, для которых u лежит в любом конечном интервале,

$$P\{\mu_r(T_n^{(2)}) = k\} = \frac{1}{\sigma_r\sqrt{2\pi n}}\, e^{-u^2/2}(1+o(1)).$$

Для изучения поведения $\eta(T_n)$ заметим, что из (4), (7) и леммы 2.2.2 [1] следует, что распределение $\eta(T_n)$ совпадает с распределением максимального заполнения ячеек в обобщенной схеме размещения n частиц по n+1 ячейке, которой соответствует последовательность независимых одинаково распределенных случайных величин $\xi_1^{(\theta)}$, $\xi_2^{(\theta)}$, ..., имеющих распределение (3). Подробно такая схема размещения рассмотрена в [1]. Для $T_n^{(1)}$ эта схема сводится к классической задаче равновероятного размещения, поэтому для $\eta(T_n^{(1)})$ справедлива теорема 2.5.2 из [1]. Аналогично для $T_n^{(2)}$ мы приходим к схеме размещения n одинаковых частиц по

$n+1$ разным ячейкам. Отсюда следует теорема 2 [2] о предельном поведении $\eta(T_n^{(2)})$. Результат для $\eta(T_n)$ будет получен в другой работе автора.

Из (4) и (7) следует, что для получения результатов о случайных величинах $\mu(t,T_n)$ и $\tau(T_n)$ достаточно воспользоваться доказанными в [1] теоремами о предельном поведении $\mu(t,G_\theta)$ и $\tau(G_\theta)$ при условии $\nu(G_\theta)=n+1$. Отсюда легко видеть, что если распределение (3) имеет максимальный шаг d, то в силу теорем 2.4.3, 2.3.2, 2.3.3 и 2.4.5 из [1] справедливы следующие утверждения.

Теорема 2. При целых $s\to\infty$ и $n=sd$ для любого фиксированного $x>0$

$$P\{(B_\theta/n)^{1/2}\tau(T_n)\leqslant x\}=\sum_{k=-\infty}^{\infty}(1-k^2x^2)e^{-k^2x^2/2}(1+o(1)).$$

Теорема 3. Для любых фиксированных целых положительных $k_1,\dots,k_t$, кратных d, при $n=sd\to\infty$

$$P\{\mu(1,T_n)=k_1,\dots,\mu(t,T_n)=k_t\}\to k_t P\{\xi_1^{(\theta)}=k_1\}\times$$

$$\times P\{\xi_1^{(\theta)}+\dots+\xi_{k_1}^{(\theta)}=k_2\}\dots P\{\xi_1^{(\theta)}+\dots+\xi_{k_{t-1}}^{(\theta)}=k_t\}.$$

Теорема 4. При $n,t\to\infty$, $t/\sqrt{n}\to 0$ для любого фиксированного $x>0$

$$P\{2\mu(t,T_n)/B_\theta t\leqslant x\}\to 1-e^{-x}-xe^{-x}.$$

Теорема 5. Если $n,t\to\infty$ так, что $t(B_\theta/n)^{1/2}\to\beta$, где β - положительная постоянная, то при любых фиксированных $x_1<x_2$

$$P\{x_1\leqslant 2\mu(t,T_n)/B_\theta t\leqslant x_2\}\to F(x_1,x_2),\ \text{где}$$

$$F(x_1,x_2)=\int_{x_1}^{x_2}\int_0^1\frac{x}{(1-y)^{3/2}}\exp\left\{-\frac{x^2\beta^2}{8(1-y)}\right\}dG_\theta(x,y),$$

и функция распределения $G_\theta(x,y)$ имеет характеристическую функцию

$$\varphi(z_1,z_2)=\left(\frac{\operatorname{sh}(\beta\sqrt{-2iz_2})}{\beta\sqrt{-2iz_2}}-iz_1\left(\frac{\operatorname{sh}(\beta\sqrt{-iz_2/2})}{\beta\sqrt{-iz_2/2}}\right)^2\right)^{-1}.$$

Очевидно, что из теорем 2-5 следуют соответствующие утверждения для $T_n^{(1)}$ и $T_{n,\varkappa}^{(2)}$, полученные в [1], а также доказанные в [2] результаты для $T_n^{(2)}$ и $T_{n,\varkappa}^{(2)}$.

ЛИТЕРАТУРА

1. Колчин В.Ф. Случайные отображения. М.: Наука, 1984.
2. Павлов Ю.Л. Некоторые свойства плоских деревьев с висячим корнем //Дискретная математика. 1992. Т.4. Вып.2. С.61-65.
3. Павлов Ю.Л. Предельные распределения высоты случайного леса// Теория вероятностей и ее применения. 1983. Т.28. Вып.3. С.449-457.
4. Земляченко В.Н., Павлов Ю.Л. Леса из плоских посаженных деревьев и ветвящиеся процессы // Труды Петрозаводского университета. Серия прикладная математика и информатика. 1992. Вып.1. С.130-135.
5. Meir A., Moon J.W. On the altitude of nodes in random trees // Can. J. Math.. 1978. V.30. N5. P.997-1015.