

# 分岐パターンのもつ階層性の解析

Hierarchical Analysis of Branching Patterns

2011 年 2 月

早稲田大学大学院 先進理工学研究科

物理学及応用物理学専攻 非平衡系物理学研究

山本 健

# 目次

第1章	背景 — 分岐パターン・Horton-Strahler 解析とその周辺	1
1.1	分岐パターンとその階層性	1
1.2	分岐パターンのモデル化 – 木・二分木	2
1.3	分岐パターンの次数付け	3
1.4	分岐比と位相的自己相似性	5
1.5	ランダムモデル	6
1.5.1	位相の異なる二分木	6
1.5.2	二分木の数え上げ – Catalan 数	8
1.5.3	ランダムモデルの定義と性質	10
1.6	本論文の構成	11
第2章	確率変数 $S_{r,n}$ のモーメント — 二分木の逐次付加による計算	13
2.1	2 次の枝の解析	13
2.2	厳密な計算の例	15
2.2.1	$S_{2,n}$ のモーメントの表式	15
2.2.2	超幾何関数の高階微分	19
2.2.3	モーメント	20
2.2.4	歪度と尖度	23
2.3	高次の枝の解析	24
2.4	$S_{r,n}$ のモーメントの漸近形と位相的自己相似性	29
第3章	期待値の漸近展開公式およびランダムモデルの位相的自己相似性	35
3.1	一般的な確率変数について	35
3.2	二分木の粗視化	35
3.3	期待値 $E[f(S_{r,n})]$ の漸近挙動	40
3.4	位相的自己相似性とその一般化	43
第4章	ランダムモデルの数値計算法	46
4.1	数値計算の必要性	46

4.2	Dyck 経路と Catalan 数 . . . . .	47
4.3	二分木と Dyck 列の対応 . . . . .	48
4.4	ランダムな Dyck 経路の生成 . . . . .	50
4.5	ランダムモデルの数値計算アルゴリズム . . . . .	54
<b>第 5 章</b>	<b>二分木に対する中心極限定理</b>	<b>56</b>
5.1	はじめに – 中心極限定理と誤差関数 . . . . .	56
5.2	$S_{2,n}$ に対する中心極限定理とその拡張 . . . . .	57
5.3	数値計算結果 . . . . .	59
5.4	結果の検討 . . . . .	59
<b>第 6 章</b>	<b>総括</b>	<b>63</b>
	謝辞	65
	参考文献	67

# 第1章 背景 — 分岐パターン・ Horton-Strahler 解析とその周辺

## 1.1 分岐パターンとその階層性

分岐・枝分かれから構成されるパターンは系の種類・サイズを問わず幅広く存在する [1–4]。生物系にみられるパターンには神経細胞、血管網、バクテリアコロニー、植物の根・枝、葉脈などがある。無生物系に現れる分岐パターンには河川網、山脈の稜線、粘性突起、樹枝状結晶、ひび割れなどがある。道路・鉄道網といった社会現象に関連した分岐パターンも多く存在する。実際に存在するパターン以外にも、計算機科学で扱われるある種のデータ構造や生物の系統樹、組織図、スポーツのトーナメント表など、思考・分類の図的表現手段としても分岐構造は頻繁に利用される。

分岐パターンは、その生成過程に応じて 3 つのタイプに分類できる [5]。

### 1. 太い枝から細い枝が出ていくもの

樹木やひび割れなど、成長する分岐パターンの多くがこのタイプに属する。また、粘性突起・樹枝状結晶の成長・DLA のように線状に伸びる枝の不安定化によって分岐が生じるものもある。

### 2. 細い枝が合併して太い枝になるもの

このタイプの典型が河川網である。自然発生的な道路網やある種のひび割れにもこのタイプに分類できるものがある。

### 3. 網目状のネットワークからいくつかの線が消えることで分岐パターンが形成されるもの

この機構によって発生する分岐パターンは少数であり、毛細血管網が典型例である。また、与えられた重み付きのネットワークに対して最小全域木を求めるアルゴリズムはこれに関連している。

分岐パターンは階層的な性質をもっている。例えば、水理学ではある河川網（水系）において河川長さ、河道の断面積、流量等が最も大きいものを本川（ほんせん）・幹川（かんせん）と呼び、幹川に合流する河川を 1 次支川（しせん）、1 次支川に合流する河川を 2 次支

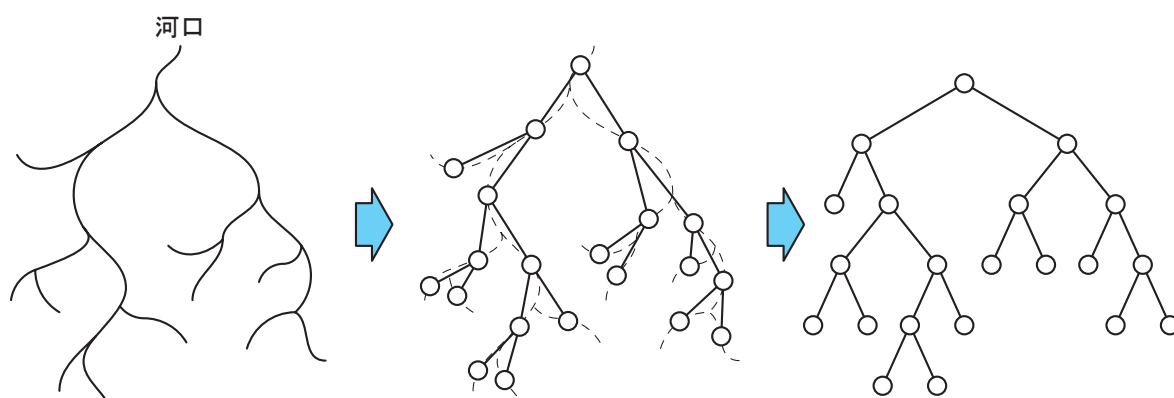


図 1.1 分岐パターンのグラフ化。(a) 河川網。(b) それぞれの流路がグラフの頂点となる。各頂点から流入している頂点へ向けて辺で結ぶ。(c) 対応する二分木。

川、… と呼び、階層性を導入している [6, 7]。なお、本川に合流する河川を支川、支川に合流する河川を小支川、以下、小小支川などと区分することもある [8]。この定義は大雑把に言えば、細い河川（重要度の低い河川）ほど大きな“次数”が割り当てられるということである（後で定義する Horton-Strahler の次数では重要度の低い流路ほど小さな次数を与える）。

分岐パターンのもつ階層性はパターンを構成する枝同士のつながり方（位相的性質）によって決まり、枝の長さや幅、分岐の角度や空間充填率といった幾何的な性質とは無関係である。本研究では分岐パターンのもつ階層性を定量化する方法の 1 つである Horton 数の数理的性質を調べる。

## 1.2 分岐パターンのモデル化 — 木・二分木

分岐パターンにおいて、パターンを構成しているセグメントを頂点とし、つながり方に対応する辺を引くとグラフが得られる（図 1.1）。このグラフは元の分岐パターンの位相構造を保持している。本研究で想定している分岐パターンはループをもたず、分岐が二股であるものに限る。このような分岐パターンから得られるグラフは二分木と呼ばれる。

グラフ理論 [9] の用語で必要なものを列挙する。まず、木とは閉路をもたない連結な無向グラフを指す。また、木のうちの 1 つの頂点を選び“根”としたものを根つき木と呼ぶ。根つき木は、各辺に“根に向かう方向”が与えられた有向グラフであるともいえる。根つき木の構造は（単為生殖の）生物の家系図や植物に似ているため、グラフ理論では根つき木に関する用語をこれらの分野から借用している。ここでは、以下で必要となる語のみを説明する。まず、辺で結ばれている 2 つの頂点  $v_1, v_2$  を親子と呼び、 $v_1$  の方が  $v_2$  より根に近い

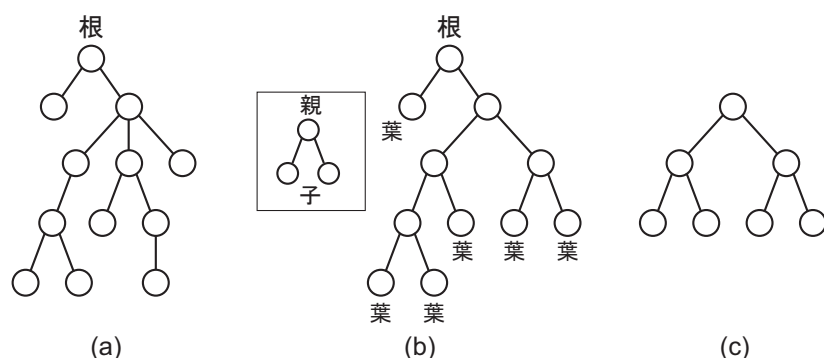


図 1.2 (a) 二分木でない根つき木、(b) 二分木 (サイズは 6)、(c) 2-完全二分木。本論文内の全ての二分木の図は根を最も上に、子頂点を親頂点の下に描かれている。

いとき  $v_1$  は  $v_2$  の親、 $v_2$  は  $v_1$  の子という。子をもたない頂点を葉と呼ぶ。根つき木の葉の総数を本論文ではサイズと呼ぶ。根つき木の特別なクラスとして二分木がある。二分木とは葉でない各頂点の子の数がつねに 2 であるような根つき木である (全く同様に  $n$  分木が定義できる)。計算機科学では、二分木は二分探索木、バイナリヒープ、構文木などに関連した重要なデータ構造である [10]。全ての葉が根から距離  $r$  だけ離れているような二分木を  $r$ -完全二分木と呼ぶ。 $r$ -完全二分木のサイズは  $2^r$  である。本節で挙げた用語は図 1.2 にまとめてある。

### 1.3 分岐パターンの次数付け

分岐パターンの階層性に関する研究は Horton [11] による河川網の分岐の解析から始まった。この解析では、まず河川網を構成する各々の流路の重要度を反映した次数を計算する (重要度は源流に近いほど小さく、河口に近いほど大きい)。分岐パターンをグラフ化した二分木においては頂点に次数を割り当てることに対応する。次数が与えられた二分木において、次数  $r$  をもつ頂点からなる最大の路を  $r$  次の枝と呼ぶ。

Horton による二分木の次数付けは以下のルールからなる。

- (1) 二分木の葉を全て次数 1 とする。
- (2) 同じ次数  $r$  の枝の合流が起こったら、その合流点の次数を  $r+1$  とする。さらに、合流する次数  $r$  の 2 本の枝のうち長い方を次数  $r+1$  に書き換える。2 本の枝の長さが同じであるときはどちらか一方のみを選んで次数を  $r+1$  にする。
- (3) 異なる次数の枝の合流が起こったら、大きい方の次数を合流点の次数とする。

このルールを繰り返し適用し、二分木の全ての頂点に次数を割り当てる。河川網に対する本来の Horton のルールでは (2) においてどちらの枝の次数の書き換えを行なうかを河川の長さや合流の角度（幾何学的な情報）に基づいて決定している。このため Horton のルールは純粹に位相的ではない。上のルール (1)-(3) は位相的性質のみから次数を計算できるように修正を加えたものである。Horton のルールでは次数の書き換えによって 1 つの頂点の次数が何回も更新され、無駄が多いといえる。また、枝の長さを逐次計算する必要があるという点も効率が悪い。

これらの難点を取り払ったのが Strahler による次数付け [12,13] である。

- (1) 二分木の葉を全て次数 1 とする。
- (2) 同じ次数  $r$  の枝の合流が起こったら、合流点の次数を  $r + 1$  とする。
- (3) 異なる次数の合流が起こったら、大きい方の次数を合流点の次数とする。

このルールは Horton のものと似ているが、同じ次数の枝の合流が起こった場合の処理が簡単になっている。Strahler のルールでは枝の長さの計算・次数の書き換えが不要であり、各頂点の次数は 2 つの子の次数のみにより局所的に決定できる。Strahler のルールの欠点は異なる次数の合流が起こった際に次数の小さい方の枝の情報が捨てられてしまうということである。

Strahler のルールがもつ欠点を補ったものとして Shreve による次数付け [15] がある。

- (1) 二分木の葉を全て次数 1 とする。
- (2) 葉以外の頂点の次数は 2 つの子の次数の和とする。

このルールでは、二分木の根がもつ次数はその二分木の葉の個数に等しい。合流点での次数の増加が一定でないため、どの次数の枝が存在するのか分かりにくいという難点がある。

Strahler のルールを拡張した手法として、2 種類の次数を利用するものがある [14]。この手法では、まず Strahler の次数付けを行い、頂点の次数と親の次数を組にして二重次数とする。二重次数はパターンのフラクタル次元とも強い関連があるが、2 つの数を取り扱うため、理論的な取り扱いは比較的難しい。

本節で挙げた 4 つの次数付けの適用例を図 1.3 に示す。ルールの簡便さおよび理論計算の見通しの面から、本論文では Strahler のルールによる次数のみを取り扱い、単に“次数”といえは Strahler の次数を指すものとする。Strahler の次数を用いた二分木の解析を Horton-Strahler 解析ともいう。本節に挙げたもの以外にも次数付けのルールが提案されている [16-18]。

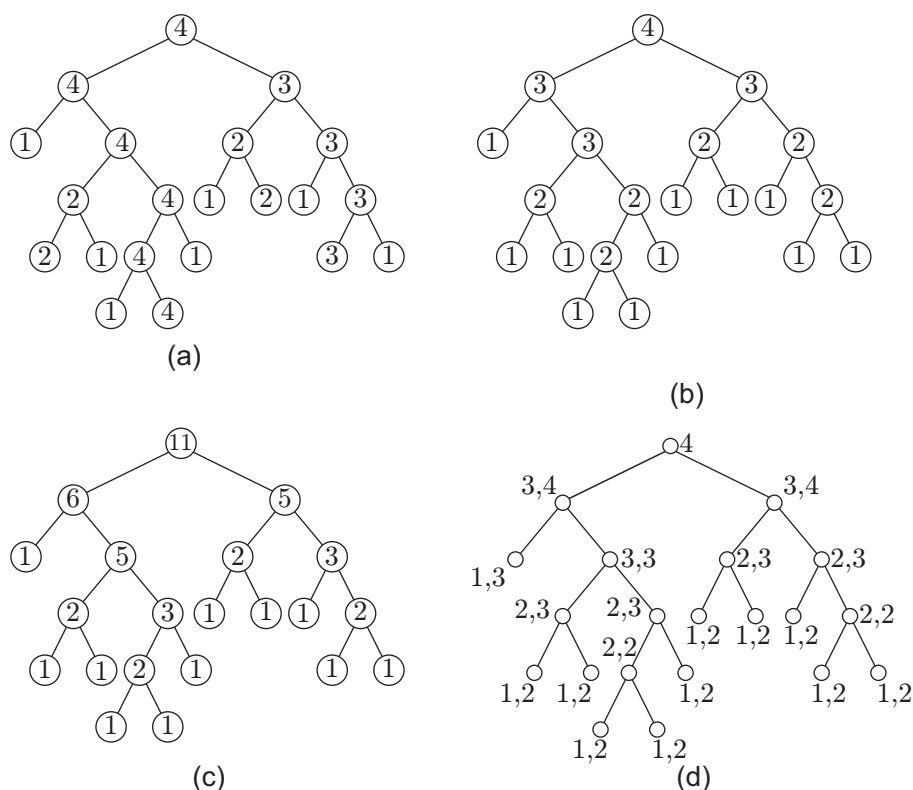


図 1.3 次数付けの例。(a) Horton のルール、(b) Strahler のルール、(c) Shreve のルール、(d) 二重次数。

### 1.4 分岐比と位相的自己相似性

Horton-Strahler 解析では次数  $r$  の枝の本数が重要である（次数  $r$  の頂点の数はそれほど重要でない）。Strahler の次数付けにおいて、次数  $r$  の枝が現れるには次数  $r - 1$  の枝同士の合流が必要である。 $r + 1$  次の枝 1 本あたりの  $r$  次の枝の本数を  $r$  次の分岐比と呼ぶ [11]。大まかにいえば、分岐比はパターンの分岐の複雑さを表している。 $r$  次の分岐比が大きければ、 $r$  次の枝が“側枝”（より高次の枝に流入している枝）として多く存在する。特に 完全二分木では分岐比は次数によらず 2 である。この分岐比 2 は分岐比の最小値でもある（分岐比 2 は同じ次数の枝同士が無駄なく合流する場合にのみ実現できる）。

パターンの一部が全体と相似になっている性質を自己相似性といい、自己相似性をもつパターン・図形をフラクタルと呼ぶ [19]。フラクタルパターンの空間的な分布を表す尺度としてフラクタル次元がある。多くの分岐パターンも自己相似性をもつことが調べられており、アマゾン川のフラクタル次元は 1.85、ナイル川は 1.40 [20]、植物では枝葉のフラクタル次元  $1.28 - 1.79$  [21]、根のフラクタル次元  $1.48 - 1.58$  [22] といった測定も数多く行な



われている。Mandelbrot [19] は河川網が地表面を覆うような形態（平面充填曲線に近い）であるために理想的にはフラクタル次元が 2 となることを指摘している（河川網の典型的なフラクタル次元は 1.5 から 2 の値である [23, 24]）。

自己相似性やフラクタル次元はパターンのもつ幾何学的性質と関連している。そのため、分岐パターンの位相的性質のみを抽出して得られる二分木にこれらの概念を適用しても意味がないように思われる。しかし、自己相似性はパターンの解析において重要であるので、類似の性質を位相的な構造だけを用いて構成するのも重要である。

実際の分岐パターンでは次数が大きい枝ほど長く太くなることが多い。例えば、河川では下流ほど川幅が大きくなり、植物では幹に近いほど枝が太くなる。つまり、大雑把に言えば次数は枝の空間的なスケールを表している。分岐比については、低次の分岐比は小さなスケールでの分岐の様子を、高次の分岐比は大きなスケールでの分岐の様子を反映していると考えられる。そして、分岐比が次数によらず一定であればある種の自己相似性（分岐の様子がスケールに依存しない）をもっているといえるだろう。

以上の観察を踏まえ、分岐パターンの分岐比が次数によらず（ほぼ）一定であるとき、その分岐パターンは位相的自己相似性をもつという [25]。分岐比は位相的な量であったので、位相的自己相似性も文字通り位相的性質である。なお、第3章で導入する“二分木の粗視化”を利用すると、位相的自己相似性と元々の自己相似性の類似をより深く理解できる。表 1.1 に位相的自己相似性をもつ分岐パターンの例とそれらの分岐比を挙げる。

歴史的には河川網の位相的自己相似性が調べられたのが Horton-Strahler 解析の出発点であった [11]。多くの河川網は位相的に自己相似であり、その分岐比は 3 から 5 程度の値をとる [26]。河川網の位相的自己相似性を特に Horton の（第 1）法則と呼ぶ（そのほかに流路の長さに関する第 2 法則、流路の勾配に関する第 3 法則、流域面積に関する第 4 法則がある）[27]。

## 1.5 ランダムモデル

### 1.5.1 位相の異なる二分木

本研究では“与えられたサイズをもつ二分木の全体”といった集合を取り扱う。そのために、まず 2 つの二分木が“同じ”であるかどうかの基準を与える。

まず、2 つの二分木  $T, T'$  が同型であるとは  $T$  の頂点集合  $V$  から  $T'$  の頂点集合  $V'$  への親子関係を保つ全単射  $\phi: V \rightarrow V'$  が存在することと定義される（グラフの同型の一般的な定義 [41] を二分木に焼き直した）。“親子関係を保つ”とは、 $v^* \in V$  が  $v \in V$  の子ならば  $\phi(v^*) \in V'$  は  $\phi(v) \in V'$  の子となることである。定義より、 $\phi$  は  $T$  の根を  $T'$  の根にうつし、 $T$  の葉を  $T'$  の葉にうつす。このため、 $T$  と  $T'$  の葉の数が異なれば上のような写

表 1.1 位相的に自己相似な分岐パターンとその分岐比の例。

分岐系	内容	分岐比
河川	マロニー川 [28]	4.16
	ケンタッキー川 [29]	4.6
	バッククリーク川 [30]	4.1
	ケンタッキー川 [30]	3.95
植物	モミノキ [31]	4.8
	トネリコ [31]	6.5
	ユリノキ [31]	4.7
	サンザシ [32]	6.92
	アメリカクロモジ [32]	3.80
	ガマズミ [32]	3.80
サンゴ [33]	<i>L. hebes</i>	3.7, 3.9
	<i>L. virgulata</i>	3.1
カイメン [34]	<i>Raspailia inaequalis</i>	$2.71 \pm 0.046$
血管	ヒトの肺 [35]	3.15
	ブタの心臓 [36]	3.30
亀裂 [37]	ジメタクリレート樹脂	$2.5 \pm 0.05$
林道 [38]	飯能	2.33
	奥多摩	2.66
	諸塚	2.20
DLA [39]		5.15
ネットワーク [40]	コミュニティ構造の木	3.46

像  $\phi$  を構成することができない。ゆえに、サイズが異なる二分木は同型になりえない。図 1.4 に示した二分木のうち  $T_2, T_3$  および  $T_4$  は同型であるが  $T_1$  はこれらとは同型でない。

二分木  $T$  と  $T'$  の位相が等しいとは、

- (1)  $T$  と  $T'$  は同型
- (2) 同型  $\phi: T \rightarrow T'$  は左右を保つ ( $T$  の頂点  $v_L$  が  $v$  の左側の子ならば、頂点  $\phi(v_L)$  は  $\phi(v)$  の左側の子である。また、 $T$  の頂点  $v_R$  が  $v$  の右側の子ならば、頂点  $\phi(v_R)$  は  $\phi(v)$  の右側の子である)

を満たすことと定義する。この概念は同型性よりもきめ細かく、図 1.4 の  $T_2$  と  $T_3$  は同型であったが位相は異なる。一般に、ある二分木とその鏡像は位相が等しいとは限らない(子

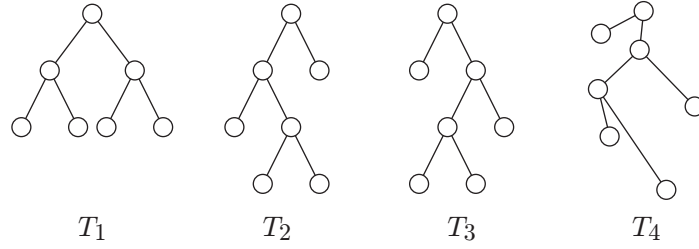


図 1.4 3 つの二分木  $T_2, T_3$ , および  $T_4$  は同型であるが、 $T_1$  はこれらとは同型でない。二分木  $T_2$  と  $T_3$  は互いに鏡像の関係にあるが位相は異なる。二分木  $T_3$  と  $T_4$  は位相が等しい (グラフ理論では辺の長さや角度の情報は無視してよい)。

の左右を区別しているため)。同型でないグラフは位相が等しくなりえないので、サイズが異なる二分木はつねに位相が異なる。

### 1.5.2 二分木の数え上げ – Catalan 数

サイズ  $n$  の位相が異なる二分木の全体を  $\Omega_n$  で表し、 $\Omega_n$  の要素の数を  $c_{n-1}$  とする。具体的に書き出すことで

$$\#\Omega_1 = c_0 = 1, \quad \#\Omega_2 = c_1 = 1, \quad \#\Omega_3 = c_2 = 2, \quad \#\Omega_4 = c_3 = 5, \dots$$

が確認できる (図 1.5)。ただし、“ $\#$ ” は有限集合の元の個数を与える記号とする。一般の  $c_{n-1}$  の表式を求める方法がいくつか知られているが、本節では母関数を利用する方法 [42–45] を紹介する。

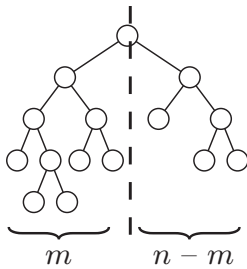
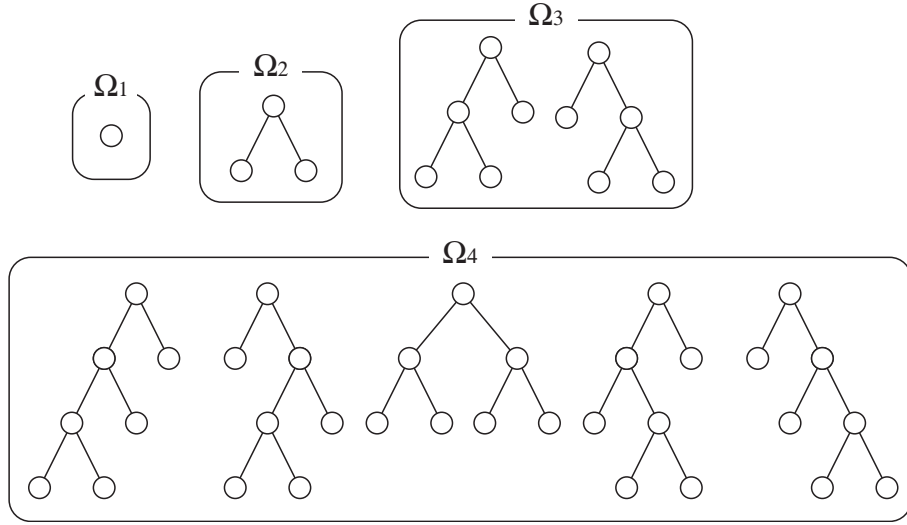


図 1.6 二分木を二分する。

サイズ  $n$  の二分木の根を取り除くと左右 2 つの二分木 (元の二分木の部分木) に分かれる (図 1.6)。左側の部分木のサイズが  $m$  であるとすれば右側の部分木のサイズは  $n - m$  となる (2 つの二分木のサイズの合計が  $n$  であるため)。左側の部分木のとりうるパターンの総数は  $c_{m-1}$  である。同様に右側の部分木のパターンは  $c_{n-m-1}$  とおりである。したがって、このように分かれる場合の異なる木は全部で  $c_{m-1}c_{n-m-1}$  とおりある。 $m$  は 1 から  $n - 1$  までの値をとりうるので、可能なパターンを全て数え上げると

$$c_{n-1} = \sum_{m=1}^{n-1} c_{m-1}c_{n-m-1} \quad (1.1)$$

という関係が分かる。右辺はたたみ込みの形になっている。

図 1.5 異なる位相をもつ二分木の全体  $\Omega_n$  ( $n = 1, 2, 3, 4$ )。

数列  $\{c_{n-1}\}_{n=1,2,\dots}$  の母関数を

$$G(x) = \sum_{n=1}^{\infty} c_{n-1} x^n \quad (1.2)$$

とおく。たたみ込み (1.1) を意識して  $G(x)$  の 2 乗を計算すると

$$\begin{aligned} G(x)^2 &= \left( \sum_{n=1}^{\infty} c_{n-1} x^n \right) \left( \sum_{m=1}^{\infty} c_{m-1} x^m \right) \\ &= \sum_{n=2}^{\infty} \left( \sum_{m=1}^{n-1} c_{m-1} c_{n-m-1} \right) x^n \\ &= \sum_{n=2}^{\infty} c_{n-1} x^n \\ &= G(x) - c_0 x = G(x) - x. \end{aligned}$$

これは  $G(x)$  に関する 2 次方程式であり、その解は

$$G(x) = \frac{1 \pm \sqrt{1-4x}}{2}$$

である。母関数の定義 (1.2) において  $x = 0$  を代入すると  $G(0) = 0$  が得られる。これと矛盾しないように  $\pm$  のうち  $-$  (マイナス) を選び、

$$G(x) = \frac{1 - \sqrt{1-4x}}{2}$$

が母関数の陽な表示である。右边を  $x$  のベキ級数に展開し、(1.2) と比較すれば

$$c_n = \frac{(2n)!}{n!(n+1)!} = \frac{1}{n+1} \binom{2n}{n} \quad (1.3)$$

が得られる。ただし、 $\binom{n}{m}$  は二項係数である。

$c_n$  は  $n$  番目の Catalan 数と呼ばれており、組合せ論の様々な場面で現れる [46, 47]。本論文では第4章において、ある種の格子路 (Dyck 経路) の総数が Catalan 数で表されることを利用した数値計算アルゴリズムについて述べる。

### 1.5.3 ランダムモデルの定義と性質

ランダムモデルは Shreve [48] により導入された確率モデルであり、サイズ  $n$  の位相が異なる全ての二分木のなす集合  $\Omega_n$  に一様確率測度  $P_n$  を与えた有限確率空間  $(\Omega_n, P_n)$  のことを指す。すなわち、ランダムモデル  $(\Omega_n, P_n)$  では任意の二分木  $T \in \Omega_n$  の統計的な重みが  $P_n(T) = 1/c_{n-1}$  となる。ランダムモデルにおいて重要な確率変数が  $S_{r,n} : \Omega_n \rightarrow \mathbb{N} \cup \{0\}$  である。二分木  $T \in \Omega_n$  に対して、 $S_{r,n}(T)$  は  $T$  に含まれる  $r$  次の枝の本数を与える。 $P_n$  に関する  $\Omega_n$  上の平均を  $E[\cdot]$  で表す。

ランダムモデルは様々な量を厳密に計算することが可能であるのが利点である。特に、2 次の枝の本数を与える  $S_{2,n}$  の確率論的な性質はよく調べられている [49–56]。例えば、 $S_{2,n}$  の平均および分散は

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

$$\text{var}(S_{2,n}) = \frac{n(n-1)(n-2)(n-3)}{2(2n-3)^2(2n-5)}$$

という簡単な形で表せる [49]。定義より全ての二分木  $T \in \Omega_n$  に対して  $S_{1,n}(T) = n$  という定数値となり (このことを簡単に  $S_{1,n} \equiv n$  と表す)、 $S_{1,n}$  の確率変数としての性質には全く面白みがない。したがって、確率変数  $S_{2,n}$  は “性質が自明でない確率変数のうちで最も扱いやすい” といえる。実際、 $S_{r,n}$  は次数  $r$  が大きくなるほど多くの枝の影響を受けることになり、解析的な取り扱いが複雑になる。実は、 $S_{2,n}$  の平均や分散が厳密に (しかも簡単な形で) 計算できることの背後には  $S_{2,n}$  のモーメントが超幾何関数で表されるという性質がある (第2章で示す)。一方、 $S_{3,n}, S_{4,n}, \dots$  を厳密に求めるためには “超幾何関数の和” の計算を実行する必要があるが、一般にこの計算は既知の特殊関数では表すことができない。実際、一般の  $S_{r,n}$  に関する厳密な結果はほとんど知られていない。 $S_{r,n}$  の平均お

よび分散の漸近的な性質

$$\frac{E[S_{r,n}]}{n} \rightarrow \frac{1}{4^{r-1}}, \quad \frac{\text{var}(S_{r,n})}{n} \rightarrow \frac{1}{3} 16^{1-r} (1 - 4^{1-r}) \quad \text{as } n \rightarrow \infty \quad (1.4)$$

が求められているくらいである [50]。

ランダムモデルは単純化された“数学モデル”である。実際の分岐パターンはランダムモデルにしたがって形成されるのではなく、パターンの出現しやすさは一様ではない。また、ランダムモデルではパターンの時間発展（特に成長）を記述することはできない。しかし、ランダムモデルは分岐パターンを生む系の1つの雛型であるとみなせる。ランダムモデルに対して得られた結果の多くは定性的には実際の分岐パターンにも適用可能であると考えられる。例えば、現実の多くの分岐パターンがもつ位相的自己相似性はランダムモデルにも現れる。さらにランダムモデルの平均的な分岐比はほぼ4であり、実際の分岐系の分岐比の値（表1.1）とも近い。

計算機科学の分野ではコンピュータのレジスタの割り当てに関連して二分木の根の次数の性質が調べられている [57–62]。ランダムモデル  $(\Omega_n, P_n)$  において二分木  $T \in \Omega_n$  の根の次数を  $S^{(n)}(T)$  とすると、 $E[S^{(n)}] = \log_4 n + O(1)$  が知られている [57]（この式はランダムモデルの分岐比が漸的に4であることと関係がありそうである）。また、 $S^{(n)}(T) = r$  となる二分木  $T$  の総数は

$$\#\{T \in \Omega_n | S^{(n)}(T) = r\} = \frac{4^n}{2^{r+1}} \sum_{j=1}^{2^r-1} (-1)^{j+1} \sin^2 \left( \frac{j\pi}{2^{r+1}} \right) \left[ \cos \left( \frac{j\pi}{2^{r+1}} \right) \right]^{2n-2}$$

のように煩雑だが厳密に書き表すことができる [58]。

異なる位相をもつ二分木の全体に一樣な確率を与えたのがランダムモデルであるが、サイズが一定の互いに同型でない二分木の全体に一樣確率測度を与えるモデルも考えられる。このような確率モデルは“ambilateral tree”と呼ばれる。ambilateral tree に対する Horton-Strahler 解析も行なわれているが [50, 51]、ランダムモデルよりも計算が難しい。例えば、サイズ  $n$  の ambilateral tree の総数を  $w_n$ 、その母関数を  $W(x) = \sum_{n=0}^{\infty} w_n x^n$  と表すと、母関数を決める関数方程式

$$W(x) = x + \frac{1}{2}W(x)^2 + \frac{1}{2}W(x^2)$$

が得られるが、 $W(x)$  について解析的に解くことはできない [63]。対照的に、ランダムモデルでは対応する量が Catalan 数 (1.3) として簡潔に表されていた。

## 1.6 本論文の構成

最も単純なランダムモデルですら一般の次数  $r$  の確率変数  $S_{r,n}$  についてはその数学的な性質の十分な理解には至っていない。そこで、本論文では確率変数  $S_{r,n}$  の性質（主として

漸近的な振る舞い)を調べる。

本論文は6つの章からなる。第2章では確率変数  $S_{r,n}$  の平均およびモーメントの計算を行なう。このためには“ $S_{r,n}(T) = m$  を満たす二分木  $T \in \Omega_n$  の総数”を求めることが必要となる。この章では完全二分木を逐次的に“ぶら下げていく手法”を導入し、二分木の数え上げを実行する(母関数などは用いず、直接的な数え上げを行なう)。確率変数  $S_{2,n}$  のモーメントは厳密な計算が可能で、Gauss の超幾何関数の微分を用いて表すことができる。一般の次数  $r$  の場合の  $S_{r,n}$  のモーメントは解析的には計算できない。代わりに次数  $r$  に関する漸化式を導出し、これを数値計算することで  $S_{r,n}$  のモーメントの漸近的な(すなわち  $n \rightarrow \infty$  での)振る舞いを把握することができる。また、この漸化式から  $S_{r,n}$  のモーメントの漸近的な形が求められる。これらの結果からランダムモデルの分岐比が漸的に4であり、サイズ  $n$  が十分に大きい場合にランダムモデルは位相的自己相似性をもつことが示される。

第3章では二分木に対して“粗視化”に相当する変換を導入し、その性質を調べる。これによって、 $f(S_{r,n})$  という一般的な形の確率変数に対しても第2章で導出した漸化式と類似の関係式が得られる。この漸化式は漸近の場合に解くことができ、その解が一般的な形の確率変数の期待値に対する漸近展開の公式を与える。また、ランダムモデルは位相的自己相似性を一般的な形の確率変数に拡張した強い性質をもつことが公式からほぼ自動的に示される。

第4章ではランダムモデルの数値計算アルゴリズムを提案する。方針は  $\Omega_n$  に含まれる全ての二分木を等確率で生成することである。これを実現するために Dyck 経路と呼ばれる格子路と二分木の間の一対一対応を構成する。そして、組合せ論および Markov 連鎖の考え方を利用して Dyck 経路を等確率で生成する方法を提案する。これらに関連する数学的な背景にも言及する。

第5章では確率変数  $S_{r,n}$  に対する中心極限定理を調べる。準備として中心極限定理および誤差関数の基本的な性質を述べた後に、 $S_{2,n}$  に対する中心極限定理(解析的な結果)を紹介する。この定理を2とおりのやり方で  $S_{r,n}$  に対する主張に拡張する。第4章で述べた数値計算アルゴリズムおよび第3章で導出した漸近展開の公式を援用し、中心極限定理が  $S_{r,n}$  についても成立することを検証する。

第6章は本論文全体の総括および展望に充てている。

## 第2章 確率変数 $S_{r,n}$ のモーメント— 二分木の逐次付加による計算

モーメントは確率変数を特徴付ける重要な指標の1つである。本章では小さな完全二分木を次々とぶら下げる手法を導入し、確率変数  $S_{r,n}$  のモーメントの性質を調べる。モーメントの厳密な計算は  $r \leq 2$  の場合にのみ可能であり、 $r \geq 3$  の場合のふるまいは数値計算から理解される。ただし、 $n \rightarrow \infty$  の漸近的な挙動は全ての  $r$  に対して調べることができ、その結果からランダムモデルに対する位相的自己相似性が得られる。

### 2.1 2 次の枝の解析

サイズ  $n$  の二分木のうち 2 次の枝を  $m$  本もつものの個数を

$$\mathcal{N}(n, m) = \#\{T \in \Omega_n | S_{2,n}(T) = m\}$$

とおく。これを用いると確率変数  $S_{2,n}$  の  $k$  次モーメントは

$$E[S_{2,n}^k] = \sum_{m=1}^{\lfloor n/2 \rfloor} m^k P_n(S_{2,n} = m) = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} m^k \mathcal{N}(n, m)$$

という形に表せる。ただし、 $\lfloor \cdot \rfloor$  は床関数（小数点以下の切り捨て）である。まず、 $\mathcal{N}(n, m)$  の具体形を求める。

2 次の枝は 2 本の 1 次の枝が合流することによってのみ生じうる。したがってある二分木  $T \in \Omega_n$  に 2 次の枝が  $m$  本あるならば、 $n$  個の葉のうち  $m$  組（ $2m$  個）が合流して 2 次の枝をつくり、残りの  $n - 2m$  個は 2 次以上の高次の枝に合流している。このことから 2 次の枝を  $m$  本もつ サイズ  $n$  の二分木は全て次のような方法で構成することができる（図 2.1 も見よ）。

- (1) サイズ  $m$  の二分木をつくる。
- (2)  $n - 2m$  個の葉（側枝）をぶら下げる。



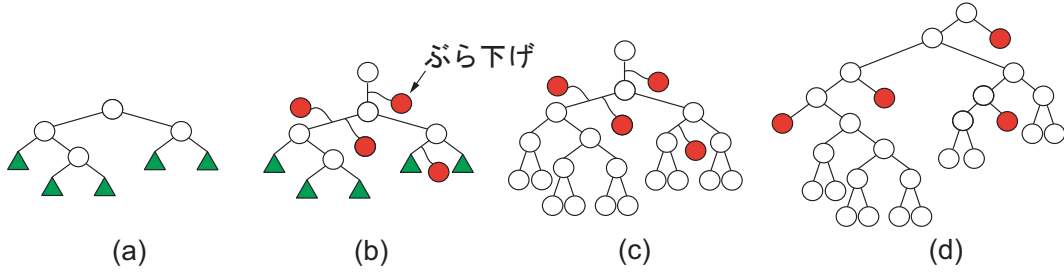


図 2.1 葉のぶら下げによる  $S_{2,n}(T) = m$  を満たす二分木  $T \in \Omega_n$  の構成法 ( $n = 14$ ,  $m = 5$  の場合)。(a) サイズ  $m$  の二分木を 1 つつくる。二分木の葉を三角形 ( $\triangle$ ) で表している。(b)  $n - 2m (= 4)$  個の葉をぶら下げる。(c) はじめの二分木の葉 (三角形の頂点) をサイズ 2 の完全二分木で置き換える。(d) 得られる二分木。

- (3) はじめの二分木の葉を全てサイズ 2 (深さ 1) の完全二分木で置き換える ( $m$  組の枝の合流に相当)。
- (4) ぶら下げた部分を頂点に置き換える。

この構成法によって 2 次の枝を  $m$  本もつ全ての二分木をつくることができ、逆に (1) の二分木または (3) の側枝の分配の仕方を変えればつくられる二分木は異なったものとなる (すなわち、過不足なく二分木をつくることができる)。したがって、上の手順の中での選択の自由度 (多重度) を数え上げることにより  $\mathcal{N}(n, m)$  を計算することができる。まず、(1) においてサイズ  $m$  の二分木の個数は  $c_{m-1}$  である。次に、(3) において  $n - 2m$  個の側枝を  $2m - 1$  本の枝 ((1) でつくった二分木の枝) へ割り当てる組合せは  $\binom{n-2}{n-2m}$  とおりある。さらに、 $n - 2m$  個の側枝は左右どちらからぶら下げるかを独立に選べることから多重度  $2^{n-2m}$  が生じる。これらの多重度は互いに独立なので、掛け合わせることで

$$\mathcal{N}(n, m) = c_{m-1} \cdot \binom{n-2}{n-2m} \cdot 2^{n-2m} = 2^{n-2m} \frac{(n-2)!}{(n-2m)!m!(m-1)!} \quad (2.1)$$

が得られる。

$\mathcal{N}(n, m)$  の具体形 (2.1) を用いると、確率変数  $S_{2,n}$  の  $k$  次モーメントは

$$E[S_{2,n}^k] = \frac{2^{n-2}}{c_{n-1}} \frac{d^k}{dt^k} e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \Big|_{t=0} \quad (2.2)$$

と計算される。ただし、 $F(\alpha, \beta; \gamma; x)$  は Gauss の超幾何関数

$$F(\alpha, \beta; \gamma; x) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{m=0}^{\infty} \frac{\Gamma(\alpha+m)\Gamma(\beta+m)}{\Gamma(\gamma+m)} \frac{x^m}{m!}$$

であり、途中の計算は次節で補完する。

$k$  次モーメントの表式 (2.2) は指数関数・超幾何関数や微分演算を含んでいて複雑だが、超幾何関数の引数  $(2-n)/2$ ,  $(3-n)/2$ , 2 が全て整数もしくは半整数であるため  $n$  を変数とする有理式になる。例えば、 $k=1$  および  $k=2$  の場合の結果を陽に書き下すと

$$E[S_{2,n}] = \frac{n(n-1)}{2(2n-3)} \quad (2.3a)$$

$$E[S_{2,n}^2] = \frac{n(n-1)(n^2-n-4)}{4(2n-3)(2n-5)} \quad (2.3b)$$

となる。これより  $S_{2,n}$  の分散は

$$\text{var}(S_{2,n}) = E[S_{2,n}^2] - E[S_{2,n}]^2 = \frac{n(n-1)(n-2)(n-3)}{2(2n-3)^2(2n-5)} \quad (2.4)$$

と計算される。(2.3a) および (2.4) は Werner [49] による結果を再現する。(2.2)-(2.4) の途中計算は次節に示す。

具体的な表示 (2.3), (2.4) から、次の漸近的な性質がただちに分かる：

$$E[S_{2,n}] \sim \frac{n}{4}, \quad E[S_{2,n}^2] \sim \frac{n^2}{16}, \quad \text{var}(S_{2,n}) \sim \frac{n}{16}.$$

2.4 節で確率変数  $S_{r,n}$  のモーメントの漸近形を導出し、より一般的な形式での漸近展開を第4章で行なう。

## 2.2 厳密な計算の例

ランダムモデルの重要な利点として、2 次の枝の本数を与える確率変数  $S_{2,n}$  については確率論的な量のいくつか厳密に計算可能なことがある。本節でそれらの計算をまとめる。なお、本節は特殊関数（主に Gauss の超幾何関数とガンマ関数）および関連した公式が多用されているため、数式の面で非常に“重たい”と思われる。

### 2.2.1 $S_{2,n}$ のモーメントの表式

まず、 $S_{2,n}$  の  $k$  次モーメント

$$E[S_{2,n}^k] = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} m^k \mathcal{N}(n, m) = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} m^k 2^{n-2m} \frac{(n-2)!}{(n-2m)!m!(m-1)!}$$

の計算からはじめる。そのために

$$G_n(x) := \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \frac{(n-2)!}{(n-2m)!m!(m-1)!} x^m$$

を導入する（確率変数  $S_{2,n}$  の確率母関数）。和の上限  $\lfloor n/2 \rfloor$  が床関数で与えられているため  $n$  の偶奇によって場合分けして計算する。具体的には、 $n$  の偶奇により

$$\left\lfloor \frac{n}{2} \right\rfloor = \begin{cases} \frac{n}{2} & (n \text{ が偶数}) \\ \frac{n-1}{2} & (n \text{ が奇数}) \end{cases}$$

という違いがある。

(i)  $n$  が偶数のとき

$n = 2p$  とおくと  $\lfloor n/2 \rfloor = p$  なので

$$\begin{aligned} G_n(x) &= \frac{1}{c_{n-1}} \sum_{m=1}^p 2^{2p-2m} \frac{(2p-2)!}{(2p-2m)!m!(m-1)!} x^m \\ &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-2m-2} \frac{(2p-2)!}{(2p-2m-2)!(m+1)!m!} x^{m+1} \\ &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-2m-2} \frac{2^{p-1}(p-1)!}{2^{p-m-1}(p-m-1)!} \frac{(2p-3)!!}{(2p-2m-3)!!} \frac{1}{(m+1)!} \frac{x^{m+1}}{m!} \end{aligned}$$

ただし、2番目の等号では  $m$  のとる範囲を1だけずらし、3番目の等号では関係式  $(2p-2)! = (2p-2)!!(2p-3)!! = 2^{p-1}(p-1)!(2p-3)!!$  を用いている。ここで、ガンマ関数の性質

$$\overbrace{\alpha(\alpha+1) \cdots (\alpha+m-1)}^{m \text{ 個}} = \frac{\Gamma(\alpha+m)}{\Gamma(\alpha)}$$

を用いると、

$$\begin{aligned} \frac{(p-1)!}{(p-m-1)!} &= \overbrace{(p-1)(p-2) \cdots (p-m)}^{m \text{ 個}} \\ &= (-1)^m (1-p)(2-p) \cdots (m-p) = (-1)^m \frac{\Gamma(1-p+m)}{\Gamma(1-p)} \end{aligned}$$

および

$$\begin{aligned} \frac{(2p-3)!!}{(2p-2m-3)!!} &= \overbrace{(2p-3)(2p-5)\cdots(2p-2m-1)}^{m \text{ 個}} \\ &= (-2)^m \left(\frac{3}{2}-p\right)\left(\frac{5}{2}-p\right)\cdots\left(\frac{2m+1}{2}-p\right) = (-2)^m \frac{\Gamma\left(\frac{3}{2}-p+m\right)}{\Gamma\left(\frac{3}{2}-p\right)} \end{aligned}$$

のように変形できる。また、階乗との関係  $(m+1)! = \Gamma(2+m)$  から

$$\frac{1}{(m+1)!} = \frac{\Gamma(2)}{\Gamma(2+m)}$$

が得られる（ガンマ関数が登場するのが超幾何関数への布石となる）。これらを代入すると

$$\begin{aligned} G_n(x) &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-m-2} (-1)^m \frac{\Gamma(1-p+m)}{\Gamma(1-p)} \cdot (-2)^m \frac{\Gamma\left(\frac{3}{2}-p+m\right)}{\Gamma\left(\frac{3}{2}-p\right)} \cdot \frac{\Gamma(2)}{\Gamma(2+m)} \frac{x^{m+1}}{m!} \\ &= \frac{2^{2p-2}}{c_{n-1}} x \frac{\Gamma(2)}{\Gamma(1-p)\Gamma\left(\frac{3}{2}-p\right)} \sum_{m=0}^{p-1} \frac{\Gamma(1-p+m)\Gamma\left(\frac{3}{2}-p+m\right)}{\Gamma(2+m)} \frac{x^m}{m!} \\ &= \frac{2^{2p-2}}{c_{n-1}} x F\left(1-p, \frac{3}{2}-p; 2; x\right) \end{aligned}$$

$p = \frac{n}{2}$  だったので、 $p$  を  $n$  に戻せば、

$$G_n(x) = \frac{2^{n-2}}{c_{n-1}} x F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; x\right)$$

が得られる。

(ii)  $n$  が奇数のとき

$n = 2p + 1$  とおくと  $\lfloor n/2 \rfloor = p$  なので

$$\begin{aligned} G_n(x) &= \frac{1}{c_{n-1}} \sum_{m=1}^p 2^{2p+1-2m} \frac{(2p+1-2)!}{(2p+1-2m)!m!(m-1)!} x^m \\ &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-2m-1} \frac{(2p-1)!}{(2p-2m-1)!(m+1)!m!} x^{m+1} \\ &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-2m-1} \frac{2^{p-1}(p-1)!}{2^{p-m-1}(p-m-1)!} \frac{(2p-1)!!}{(2p-2m-1)!!} \frac{1}{(m+1)!} \frac{x^{m+1}}{m!} \end{aligned}$$

式の変形は  $n$  が偶数のときと同様である。右辺に現れる因子のうち、

$$\frac{(p-1)!}{(p-m-1)!} = (-1)^m \frac{\Gamma(1-p+m)}{\Gamma(1-p)}$$

は上と同じであり、

$$\begin{aligned} \frac{(2p-1)!!}{(2p-2m-1)!!} &= \overbrace{(2p-1)(2p-3)\cdots(2p-2m+1)}^{m \text{ 個}} \\ &= (-2)^m \left(\frac{1}{2}-p\right)\left(\frac{3}{2}-p\right)\cdots\left(\frac{2m-1}{2}-p\right) = (-2)^m \frac{\Gamma\left(\frac{1}{2}-p+m\right)}{\Gamma\left(\frac{1}{2}-p\right)} \end{aligned}$$

と書けるので、

$$\begin{aligned} G_n(x) &= \frac{1}{c_{n-1}} \sum_{m=0}^{p-1} 2^{2p-m-1} (-1)^m \frac{\Gamma(1-p+m)}{\Gamma(1-p)} \cdot (-2)^m \frac{\Gamma\left(\frac{1}{2}-p+m\right)}{\Gamma\left(\frac{1}{2}-p\right)} \cdot \frac{\Gamma(2)}{\Gamma(2+m)} \frac{x^{m+1}}{m!} \\ &= \frac{2^{2p-2}}{c_{n-1}} x \frac{\Gamma(2)}{\Gamma\left(\frac{1}{2}-p\right)\Gamma(1-p)} \sum_{m=0}^{p-1} \frac{\Gamma\left(\frac{1}{2}-p+m\right)\Gamma(1-p+m)}{\Gamma(2+m)} \frac{x^m}{m!} \\ &= \frac{2^{2p-1}}{c_{n-1}} x F\left(\frac{1}{2}-p, 1-p; 2; x\right) \\ &= \frac{2^{n-2}}{c_{n-1}} x F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; x\right) \end{aligned}$$

結局、 $n$  の偶奇によらず

$$G_n(x) = \frac{2^{n-2}}{c_{n-1}} x F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; x\right)$$

となる。

ここで、 $x \frac{d}{dx}$  という微分演算子について

$$x \frac{d}{dx} x^m = x \cdot m x^{m-1} = m x^m$$

より、 $x \frac{d}{dx}$  の  $x^m$  への作用は  $m$  倍することと等価であることが分かる。したがって  $S_{2,n}$  の  $k$  次モーメントは

$$E[S_{2,n}^k] = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \frac{(n-2)!}{(n-2m)!m!(m-1)!} m^k = \left(x \frac{d}{dx}\right)^k G_n(x) \Big|_{x=1}$$

と表せる。さらに、形式的な変数  $x$  の変換  $x = e^t$  によって

$$x \frac{d}{dx} = x \frac{dt}{dx} \frac{d}{dt} = x \left( \frac{dx}{dt} \right)^{-1} \frac{d}{dt} = e^t e^{-t} \frac{d}{dt} = \frac{d}{dt}$$

のように書き換えられるので、

$$E[S_{2,n}^k] = \left. \frac{d^k}{dt^k} G_n(e^t) \right|_{t=0}$$

となり、(2.2) が導出された。 $t$  の関数  $M(t) := G_n(e^t)$  を微分していくことで確率変数のモーメントが順次計算されることから、確率論では  $M(t)$  はモーメント母関数（または積率母関数）と呼ばれている。

### 2.2.2 超幾何関数の高階微分

Gauss の超幾何関数に関する必要な公式をここで準備しておく。次の2つの公式 [64] は非常に有用である。

$$F(\alpha, \beta; \gamma; 1) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} \quad (\text{Gauss の公式}),$$

$$\frac{d}{dx} F(\alpha, \beta; \gamma; x) = \frac{\alpha\beta}{\gamma} F(\alpha + 1, \beta + 1; \gamma + 1; x) \quad (1 \text{ 階の微分公式}).$$

高階の微分を

$$F^{(k)}(\alpha, \beta; \gamma; x) := \frac{d^k F(\alpha, \beta; \gamma; x)}{dx^k}$$

と表す。 $F(\alpha, \beta; \gamma; x)$  を  $x$  で1回微分すると引数  $\alpha, \beta, \gamma$  がそれぞれ1ずつ増えることから、

$$\begin{aligned} F^{(k)}(\alpha, \beta; \gamma; x) &= \frac{\alpha(\alpha+1) \cdots (\alpha+k-1) \cdot \beta(\beta+1) \cdots (\beta+k-1)}{\gamma(\gamma+1) \cdots (\gamma+k-1)} F(\alpha+k, \beta+k; \gamma+k; x) \\ &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\Gamma(\beta+k)}{\Gamma(\beta)} \frac{\Gamma(\gamma)}{\Gamma(\gamma+k)} F(\alpha+k, \beta+k; \gamma+k; x). \end{aligned}$$

ここで、 $x=1$  での値は右辺に Gauss の公式を適用すれば次のように計算できる。

$$\begin{aligned} F^{(k)}(\alpha, \beta; \gamma; 1) &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\Gamma(\beta+k)}{\Gamma(\beta)} \frac{\Gamma(\gamma)}{\Gamma(\gamma+k)} \frac{\Gamma(\gamma+k)\Gamma(\gamma-\alpha-\beta-k)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \\ &= \frac{\Gamma(\alpha+k)\Gamma(\beta+k)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma-\alpha-\beta-k)}{\Gamma(\gamma-\alpha-\beta)} \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \\ &= \frac{\Gamma(\alpha+k)\Gamma(\beta+k)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma-\alpha-\beta-k)}{\Gamma(\gamma-\alpha-\beta)} F(\alpha, \beta; \gamma; 1). \end{aligned}$$

$S_{2,n}$  の性質に関連するのは  $\alpha = \frac{2-n}{2}$ ,  $\beta = \frac{3-n}{2}$ ,  $\gamma = 2$  の場合である。この場合、

$$F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{\Gamma\left(\frac{2-n}{2} + k\right) \Gamma\left(\frac{3-n}{2} + k\right) \Gamma\left(n - \frac{1}{2} - k\right)}{\Gamma\left(\frac{2-n}{2}\right) \Gamma\left(\frac{3-n}{2}\right) \Gamma\left(n - \frac{1}{2}\right)} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right)$$

となる。右辺に現れる係数は

$$\begin{aligned} \frac{\Gamma\left(\frac{2-n}{2} + k\right) \Gamma\left(\frac{3-n}{2} + k\right)}{\Gamma\left(\frac{2-n}{2}\right) \Gamma\left(\frac{3-n}{2}\right)} &= \underbrace{\frac{2-n}{2} \left(\frac{2-n}{2} + 1\right) \cdots \left(\frac{2-n}{2} + k - 1\right)}_{k \text{ 個}} \\ &\quad \times \underbrace{\frac{3-n}{2} \left(\frac{3-n}{2} + 1\right) \cdots \left(\frac{3-n}{2} + k - 1\right)}_{k \text{ 個}} \\ &= \left(-\frac{1}{2}\right)^{2k} (n-2)(n-3) \cdots (n-2k-1) \\ &= \frac{(n-2)!}{2^{2k}(n-2k-2)!} \end{aligned}$$

および

$$\begin{aligned} \frac{\Gamma\left(n - \frac{1}{2} - k\right)}{\Gamma\left(n - \frac{1}{2}\right)} &= \frac{1}{(n - \frac{1}{2} - k)(n - \frac{3}{2} - k) \cdots (n - \frac{3}{2})} \\ &= \frac{2^k}{(2n - 2k - 1)(2n - 2k + 1) \cdots (2n - 3)} \\ &= \frac{2^k(2n - 2k - 3)!!}{(2n - 3)!!} \end{aligned}$$

である。以上より高階の微分に関する公式

$$F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{1}{2^k} \frac{(n-2)!(2n-2k-3)!!}{(n-2k-2)!(2n-3)!!} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \quad (2.5)$$

が得られる。この結果は“微分が代数的な操作（掛け算）に置き換えられる”と読むことができる。

### 2.2.3 モーメント

微分公式 (2.5) を利用して  $S_{2,n}$  の低次のモーメントの計算を行なう。

まず、ほぼ自明な例として  $S_{2,n}$  の 0 次モーメントを考えてみる。明らかに  $E[S_{2,n}^0] = 1$  なので、(2.2) と比較すれば

$$F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{c_{n-1}}{2^{n-2}}$$

が得られる。この式は Gauss の公式を用いて直接計算することも可能である。

$$F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{\Gamma(2)\Gamma(2 - \frac{2-n}{2} - \frac{3-n}{2})}{\Gamma(2 - \frac{2-n}{2})\Gamma(2 - \frac{3-n}{2})} = \frac{\Gamma(n - \frac{1}{2})}{\Gamma(\frac{n+2}{2})\Gamma(\frac{n+1}{2})}.$$

分子は半整数のガンマ関数なのでただちに

$$\Gamma\left(n - \frac{1}{2}\right) = \frac{(2n-2)!}{2^{2n-2}(n-1)!}\sqrt{\pi}$$

と計算でき、分母はガンマ関数の倍数公式

$$\Gamma(2x) = \frac{2^{2x-1}}{\sqrt{\pi}}\Gamma(x)\Gamma\left(x + \frac{1}{2}\right)$$

において  $x = \frac{n+1}{2}$  とおけば

$$\Gamma\left(\frac{n+2}{2}\right)\Gamma\left(\frac{n+1}{2}\right) = \frac{\sqrt{\pi}}{2^n}\Gamma(n+1) = \frac{n!\sqrt{\pi}}{2^n}$$

と書き下せる。これらより

$$F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{(2n-2)!}{n!(n-1)!} \frac{1}{2^{n-2}} = \frac{c_{n-1}}{2^{n-2}}$$

が得られる。

次に  $S_{2,n}$  の 1 次モーメント（つまり、期待値）を計算する。(2.2) より

$$\begin{aligned} E[S_{2,n}] &= x \frac{d}{dx} G_n(x) \Big|_{x=1} = \frac{2^{n-2}}{c_{n-1}} \frac{d}{dx} \left\{ x F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; x\right) \right\} \Big|_{x=1} \\ &= \frac{2^{n-2}}{c_{n-1}} \left\{ F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + F^{(1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \right\} \end{aligned}$$

であり、微分公式 (2.5) より

$$\begin{aligned} F^{(1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) &= \frac{1}{2} \frac{(n-2)!(2n-5)!!}{(n-4)!(2n-3)!!} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\ &= \frac{(n-2)(n-3)}{2(2n-3)} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \end{aligned}$$



であるから、

$$E[S_{2,n}] = \frac{2^{n-2}}{c_{n-1}} \left\{ 1 + \frac{(n-2)(n-3)}{2(2n-3)} \right\} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) = \frac{n(n-1)}{2(2n-3)}.$$

これが  $S_{2,n}$  の期待値の厳密な表現である（この結果自体は特殊な組合せ恒等式から Werner [49] によって得られているが、一般の次数のモーメントを統一的に扱うのは不可能であった）。

2 次モーメントの計算には  $F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; x\right)$  の 2 階微分までが必要となる。微分公式 (2.5) より、

$$\begin{aligned} F^{(1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) &= \frac{(n-2)(n-3)}{2(2n-3)} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right), \\ F^{(2)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) &= \frac{(n-2)(n-3)(n-4)(n-5)}{4(2n-3)(2n-5)} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \end{aligned}$$

2 次モーメントは

$$\begin{aligned} E[S_{2,n}^2] &= \frac{d^2}{dt^2} G_n(e^t) \Big|_{t=0} = \frac{d^2}{dt^2} \left\{ e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \right\} \Big|_{t=0} \\ &= \left\{ e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) + 3e^{2t} F^{(1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \right. \\ &\quad \left. + e^{3t} F^{(2)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \right\} \Big|_{t=0} \\ &= F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + 3F^{(1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\ &\quad + F^{(2)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\ &= \frac{n(n-1)(n^2-n-4)}{4(2n-3)(2n-5)} \end{aligned}$$

と計算される。

期待値と 2 次モーメントが求まったので、 $S_{2,n}$  の分散  $\text{var}(S_{2,n})$  が計算できる：

$$\begin{aligned} \text{var}(S_{2,n}) &= E[S_{2,n}^2] - E[S_{2,n}]^2 \\ &= \frac{n(n-1)(n^2-n-4)}{4(2n-3)(2n-5)} - \left\{ \frac{n(n-1)}{2(2n-3)} \right\}^2 \\ &= \frac{n(n-1)(n-2)(n-3)}{2(2n-3)^2(2n-5)}. \end{aligned}$$

3 次以降のモーメントの計算結果を以下に挙げておく。

$$\begin{aligned} E[S_{2,n}^3] &= \frac{n(n-1)(n^4 - 2n^3 - 15n^2 + 32n + 8)}{8(2n-3)(2n-5)(2n-7)} \\ E[S_{2,n}^4] &= \frac{n(n-1)(n^6 - 3n^5 - 37n^4 + 159n^3 - 16n^2 - 424n + 144)}{16(2n-3)(2n-5)(2n-7)(2n-9)} \\ E[S_{2,n}^5] &= \frac{n(n-1)(n^8 - 4n^7 - 74n^6 + 476n^5 - 55n^4 - 4592n^3 + 8120n^2 - 992n - 1440)}{32(2n-3)(2n-5)(2n-7)(2n-9)(2n-11)} \\ &\vdots \end{aligned}$$

## 2.2.4 歪度と尖度

$S_{2,n}$  の 3 次モーメントおよび 4 次モーメントが求められたので歪度と尖度が計算できる。確率変数  $X$  の歪度  $\gamma_1$  および尖度  $\gamma_2$  は

$$\gamma_1 := \frac{E[(X - E[X])^3]}{(\text{var}(X))^{3/2}}, \quad \gamma_2 := \frac{E[(X - E[X])^4]}{(\text{var}(X))^2} - 3$$

と定義される [65]。  $\gamma_1$  の定義において  $(\text{var}(X))^{3/2}$  は“標準偏差の 3 乗”の意味である。また尖度  $\gamma_2$  については、 $X$  が正規分布にしたがうとき  $\gamma_2 = 0$  となるように  $\gamma_2$  の定義において 3 を引いている。歪度は確率分布の左右非対称性を、尖度は確率分布のピークの鋭さを表す指標である。

$S_{2,n}$  の 3 次の中心化モーメントは

$$\begin{aligned} E[(S_{2,n} - E[S_{2,n}])^3] &= E[S_{2,n}^3] - 3E[S_{2,n}^2]E[S_{2,n}] + 2E[S_{2,n}]^3 \\ &= \frac{3n(n-1)(n-2)(n-3)}{2(2n-3)^3(2n-5)(2n-7)} \end{aligned}$$

であり、4 次の中心化モーメントは

$$\begin{aligned} E[(S_{2,n} - E[S_{2,n}])^4] &= E[S_{2,n}^4] - 4E[S_{2,n}^3]E[S_{2,n}] + 6E[S_{2,n}^2]^2 - 3E[S_{2,n}]^4 \\ &= \frac{n(n-1)(n-2)(n-3)(6n^5 - 73n^4 + 320n^3 - 627n^2 + 558n - 162)}{4(2n-3)^4(2n-5)(2n-7)(2n-9)} \end{aligned}$$

である。

$S_{2,n}$  の歪度および尖度の計算結果は次のとおりである：

$$\begin{aligned} \gamma_1 &= \frac{E[(S_{2,n} - E[S_{2,n}])^3]}{\text{var}(S_{2,n})^{3/2}} = \sqrt{\frac{18(2n-5)}{(n(n-1)(n-2)(n-3)(2n-7))^2}}, \\ \gamma_2 &= \frac{E[(S_{2,n} - E[S_{2,n}])^4]}{\text{var}(S_{2,n})^2} - 3 = \frac{(2n-5)(6n^5 - 73n^4 + 320n^3 - 627n^2 + 558n - 162)}{n(n-1)(n-2)(n-3)(2n-7)(2n-9)} - 3. \end{aligned}$$

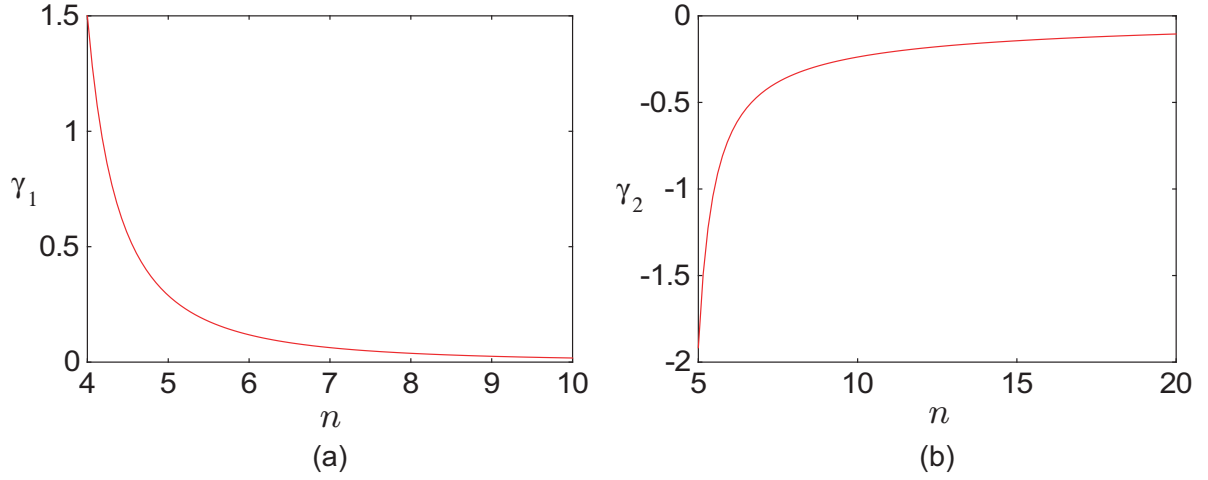


図 2.2 (a)  $S_{2,n}$  の歪度  $\gamma_1$ 、(b)  $S_{2,n}$  の尖度  $\gamma_2$  のグラフ。ともに  $n$  の関数として表示してある。

歪度および尖度のグラフを図 2.2 に示す。 $\gamma_1, \gamma_2$  は  $n \rightarrow \infty$  でともに 0 に収束する。これより、確率変数  $S_{2,n}$  の漸近的なふるまいは正規分布に近い可能性が示唆される（正規分布は尖度および歪度がともに 0 である）。より精密な表現として、第 5 章で  $S_{2,n}$  に対する中心極限定理（およびその拡張）について述べる。

## 2.3 高次の枝の解析

本節では、前節で述べた“ぶら下げ”の手法を拡張し、確率変数  $S_{r,n}$  ( $r \geq 3$ ) の  $k$  次モーメントの満たす漸化式を導出する。

サイズ  $n$  ( $\equiv n_1$ )、2 次の枝  $n_2$  本、 $\dots$ 、 $r$  次の枝  $n_r$  本 であるような二分木の総数を

$$\mathcal{N}(n, n_2, \dots, n_r) = \#\{T \in \Omega_n | S_{2,n} = n_2, \dots, S_{r,n} = n_r\}$$

とおく。このような二分木は以下のようにして構成することができる。

- (1) サイズ  $n_r$  の二分木をつくる。
- (2) その二分木の全ての葉を  $r$ -完全二分木で置き換える。
- (3)  $n_{r-1} - 2n_r$  個の  $(r-1)$ -完全二分木を次数  $r$  以上の頂点の上にぶら下げる。
- (4) 同様に、 $n_{r-2} - 2n_{r-1}$  個の  $(r-2)$ -完全二分木を次数  $r-1$  以上の頂点の上にぶら下げる。

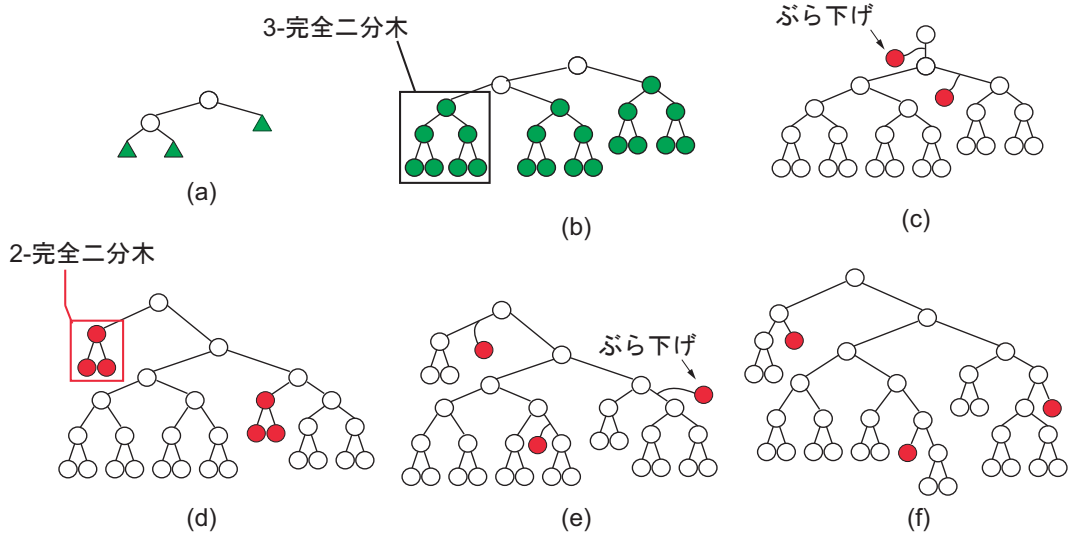


図 2.3  $n = 19, n_2 = 8, n_3 = 3$  を満たす二分木の作成手順。(a) サイズ  $n_3$  の二分木 (b) 二分木の葉を 3-完全二分木に置き換える。(c)  $n_2 - 2n_3 (= 2)$  個の葉をぶら下げる。ただし、ぶら下げられるのは次数 3 以上の頂点の上。(d) ぶら下げた葉を 2-完全二分木に置き換える。(e)  $n - 2n_2 (= 3)$  個の葉をぶら下げる。ただし、ぶら下げられるのは次数 2 以上の頂点の上。(f) 得られる二分木。

(5) この操作を繰り返す ( $n_i - 2n_{i+1}$  個の  $i$ -完全二分木を次数  $i + 1$  以上の頂点の上ぶら下げる)。

(6) 1-完全二分木まで到達したら終了。

この方法では完全二分木を階層的にぶら下げていくことで条件を満たす二分木を得ることができる。特に、 $r = 2$  の場合は前節で示した構成法と一致する。図 2.3 は  $n = 19, n_2 = 8, n_3 = 3$  である二分木を構成する手順を示している。

木の個数  $\mathcal{N}(n, n_2, \dots, n_r)$  を式で表すためには、前節と同様に構成法 (1)-(6) に現れる多重度を数え上げねばならない。まず (1) において、サイズ  $n_r$  の二分木の総数は  $c_{n_r-1}$  である。 $i$ -完全二分木をぶら下げる段階において、ぶら下げが可能な辺は  $2n_{i+1} - 1$  本ある。 $n_i - 2n_{i+1}$  個の  $i$ -完全二分木をこの  $2n_{i+1} - 1$  本の辺に分配する組合せは  $\binom{n_i-2}{n_i-2n_{i+1}}$  とおりある。各々の  $i$ -完全二分木を左右どちらからぶら下げるかを反映して多重度  $2^{n_i-2n_{i+1}}$  が生じる。これらの多重性は各段階  $i = r - 1, r - 2, \dots, 1$  ごとに独立である。したがって、全ての因子を掛け合わせることで

$$\mathcal{N}(n, n_2, \dots, n_r) = c_{n_r-1} \prod_{i=1}^{r-1} 2^{n_i-2n_{i+1}} \binom{n_i-2}{n_i-2n_{i+1}}$$

が得られる。ただし、右辺において記法の簡略化のために  $n_1 \equiv n$  としている。

特に、 $n_r = 1$  のとき

$$\mathcal{N}(n, n_2, \dots, n_{r-1}, 1) = \prod_{i=1}^{r-1} 2^{n_i - 2n_{i+1}} \binom{n_i - 2}{n_i - 2n_{i+1}}$$

は Shereve [48] によって（本研究とは異なる動機・手法にもとづいて）得られている。

$i$  次の枝は  $(i-1)$  次の枝同士の合流によってのみ生じることから、サイズ  $n$  と  $r$  次の枝の本数  $n_r$  を与えたとき  $i$  次の枝の本数  $n_i$  は  $2^{r-i}n_r \leq n_i \leq \lfloor n_{i-1}/2 \rfloor$  を満たす ( $i = 2, \dots, r-1$ )。 “ $\Omega_n$  の中で  $r$  次の枝をちょうど  $n_r$  本もつ二分木の個数” を求めるには  $(n_2, n_3, \dots, n_{r-1})$  の全ての可能な組について  $\mathcal{N}(n, n_2, \dots, n_r)$  を足し上げればよいので、

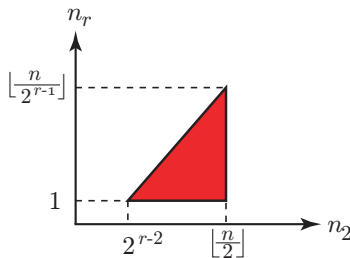
$$\#\{T \in \Omega_n | S_{r,n}(T) = n_r\} = \sum_{n_2=2^{r-2}n_r}^{\lfloor n/2 \rfloor} \sum_{n_3=2^{r-3}n_r}^{\lfloor n_2/2 \rfloor} \cdots \sum_{n_{r-1}=2n_r}^{\lfloor n_{r-2}/2 \rfloor} \mathcal{N}(n, n_2, n_3, \dots, n_r)$$

のように表せる。したがって確率変数  $S_{r,n}$  の  $k$  次モーメントは

$$\begin{aligned} E[S_{r,n}^k] &= \frac{1}{c_{n-1}} \sum_{n_r=1}^{\lfloor n/2^{r-1} \rfloor} (n_r)^k \#\{T \in \Omega_n | S_{r,n}(T) = n_r\} \\ &= \frac{1}{c_{n-1}} \sum_{n_r=1}^{\lfloor n/2^{r-1} \rfloor} \sum_{n_2=2^{r-2}n_r}^{\lfloor n/2 \rfloor} \sum_{n_3=2^{r-3}n_r}^{\lfloor n_2/2 \rfloor} \cdots \sum_{n_{r-1}=2n_r}^{\lfloor n_{r-2}/2 \rfloor} (n_r)^k c_{n_r-1} \prod_{i=1}^{r-1} 2^{n_i - 2n_{i+1}} \binom{n_i - 2}{n_i - 2n_{i+1}} \end{aligned}$$

という形で具体的に書き表せる。この式と似た表示が文献 [66] に見られるが、導出の際の数学的な厳密性・一般性は検討されていないようである。この式はきわめて複雑な形をしているが、以下に示すような“和の交換”を行なうことで相当に簡略化される。

左図より  $n_2, n_r$  に関する和は



$$\sum_{n_r=1}^{\lfloor n/2^{r-1} \rfloor} \sum_{n_2=2^{r-2}n_r}^{\lfloor n/2 \rfloor} = \sum_{n_2=2^{r-2}}^{\lfloor n/2 \rfloor} \sum_{n_r=1}^{\lfloor n_2/2^{r-2} \rfloor}$$

のように交換できる。

図 2.4  $n_2, n_r$  の和の範囲

同様に、

$$\sum_{n_r=1}^{\lfloor n_2/2^{r-2} \rfloor} \sum_{n_3=2^{r-3}n_r}^{\lfloor n_2/2 \rfloor} = \sum_{n_2=2^{r-3}}^{\lfloor n_2/2 \rfloor} \sum_{n_r=1}^{\lfloor n_3/2^{r-3} \rfloor}$$

$$\vdots$$

のように  $n_r$  に関する和を右に移動できる。これを繰り返して  $n_r$  に関する和を最も右まで移動すると、

$$\begin{aligned} E[S_{r,n}^k] &= \frac{1}{c_{n-1}} \sum_{n_2=2^{r-2}}^{\lfloor n/2 \rfloor} \sum_{n_3=2^{r-3}}^{\lfloor n_2/2 \rfloor} \cdots \sum_{n_r=1}^{\lfloor n_{r-1}/2 \rfloor} \prod_{i=1}^{r-1} 2^{n_i-2n_{i+1}} \binom{n_i-2}{n_i-2n_{i+1}} \\ &= \frac{1}{c_{n-1}} \sum_{n_2=2^{r-2}}^{\lfloor n/2 \rfloor} 2^{n-2n_2} \binom{n-2}{n-2n_2} \times \sum_{n_3=2^{r-3}}^{\lfloor n_2/2 \rfloor} 2^{n_2-2n_3} \binom{n_2-2}{n_2-2n_3} \times \cdots \\ &\quad \times \sum_{n_r=1}^{\lfloor n_{r-1}/2 \rfloor} n_r^k 2^{n_{r-1}-2n_r} \binom{n_{r-1}-2}{n_{r-1}-2n_r} \end{aligned} \quad (2.6)$$

のようになる。あまり簡単になった気配がないが、確率変数  $S_{r-1,n_2}$  の  $k$  次モーメントを計算してみると

$$E[S_{r-1,n_2}^k] = \frac{1}{c_{n_2-1}} \sum_{n_3=2^{r-3}}^{\lfloor n_2/2 \rfloor} 2^{n_2-2n_3} \binom{n_2-2}{n_2-2n_3} \times \cdots \times \sum_{n_r=1}^{\lfloor n_{r-1}/2 \rfloor} n_r^k 2^{n_{r-1}-2n_r} \binom{n_{r-1}-2}{n_{r-1}-2n_r} \quad (2.7)$$

が得られる。(2.6) と (2.7) を比較すれば

$$\begin{aligned} E[S_{r,n}^k] &= \frac{1}{c_{n-1}} \sum_{m=2^{r-2}}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} E[S_{r-1,m}^k] \\ &= \frac{1}{c_{n-1}} \sum_{m=2^{r-2}}^{\lfloor n/2 \rfloor} \frac{2^{n-2m}(n-2)!}{(n-2m)!m!(m-1)!} E[S_{r-1,m}^k] \end{aligned}$$

という関係式となる( (2.6) の  $n_3$  に関する和から後の部分は (2.7) にも共通して現れている )。

二分木のサイズ  $m$  について、 $m < 2^{r-2}$  ならば  $S_{r-1,m} \equiv 0$  (つまり  $E[S_{r-1,m}^k] = 0$ ) である。よって、右辺の和の下限を  $m = 1$  まで広げて和の値は変わらない。ゆえに

$$E[S_{r,n}^k] = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} \frac{2^{n-2m}(n-2)!}{(n-2m)!m!(m-1)!} E[S_{r-1,m}^k] \quad (2.8)$$

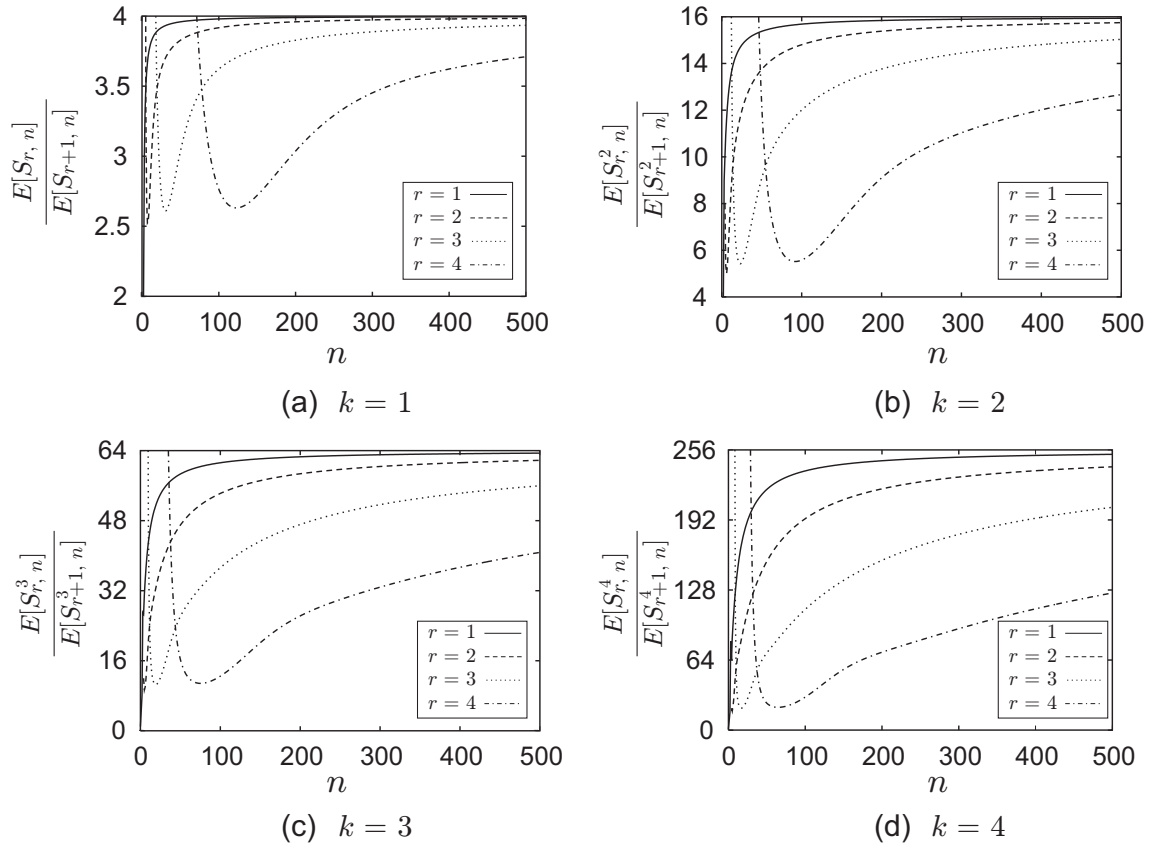


図 2.5 漸化式 (2.8) の数値計算結果。モーメントの比  $E[S_{r,n}^k] / E[S_{r+1,n}^k]$  をサイズ  $n$  の関数として表示している。

が得られる。(2.8) は各  $k$  に対して  $r$  に関する漸化式とみなすことができる。ただし、漸化式の初期条件 ( $r = 1$  のときの値) は  $E[S_{1,n}^k] = n^k$  である。例えば  $S_{3,n}$  のモーメントを計算するためには超幾何関数を含んだ項の和を計算せねばならないため、 $r \geq 3$  については  $S_{r,n}$  のモーメントは既知の特殊関数では表せないであろう。しかし、数値計算や漸近表示 ( $n \rightarrow \infty$ ) の導出は可能である。

漸化式 (2.8) を  $k = 1, 2, 3, 4$ ,  $r = 1, 2, 3, 4$  の場合に数値的に計算した結果が図 2.5 である (隣り合う次数のモーメントの比  $E[S_{r,n}^k] / E[S_{r+1,n}^k]$  をプロットしている)。  $r$  によらずに

$$\frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} \rightarrow 4^k \quad (\text{as } n \rightarrow \infty)$$

が予想され、 $k$  の値が同じなら  $r$  が大きいほど収束が遅いことが示唆される。

## 2.4 $S_{r,n}$ のモーメントの漸近形と位相的自己相似性

図 2.5 より、 $S_{r,n}$  と  $S_{r+1}$  のモーメントの比について

$$\lim_{n \rightarrow \infty} \frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} = 4^k$$

が成り立つことが示唆された。本節ではこれが正しいことを示し、さらに正確な漸近挙動（収束の速さ等）について議論する。

まず、 $F^{(\ell)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right)$  の漸近形（ $n \rightarrow \infty$  でのふるまい）を導出する。高階微分の公式 (2.5) より、

$$\begin{aligned} F^{(\ell)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) &= \frac{1}{2^\ell} \frac{(n-2)!(2n-2\ell-3)!!}{(n-2\ell-2)!(2n-3)!!} F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\ &= \frac{1}{2^\ell} \frac{(n-2)(n-3) \cdots (n-2\ell-1)}{(2n-3)(2n-5) \cdots (2n-2\ell-1)} \frac{c_{n-1}}{2^{n-2}} \end{aligned}$$

分子に現れる因子は

$$\begin{aligned} \underbrace{(n-2)(n-3) \cdots (n-2\ell-1)}_{2\ell \text{ 個}} &= n^{2\ell} - \{2+3+\cdots+(2\ell+1)\}n^{2\ell-1} + O(n^{2\ell-2}) \\ &= n^{2\ell} \left\{ 1 - \ell(2\ell+3)\frac{1}{n} + O(n^{-2}) \right\} \end{aligned}$$

と展開できる。また、分母に現れる因子は

$$\frac{1}{2n-m} = \frac{1}{2n} \left( 1 + \frac{m}{2n} + O(n^{-2}) \right)$$

を用いることで、

$$\begin{aligned} &\frac{1}{\underbrace{(2n-3)(2n-5) \cdots (2n-2\ell-1)}_{\ell \text{ 個}}} \\ &= \left( \frac{1}{2n} \right)^\ell \left( 1 + \frac{3}{2n} + O(n^{-2}) \right) \left( 1 + \frac{5}{2n} + O(n^{-2}) \right) \cdots \left( 1 + \frac{2\ell+1}{2n} + O(n^{-2}) \right) \\ &= \left( \frac{1}{2n} \right)^\ell \left\{ 1 + \frac{1}{2n} (3+5+\cdots+(2\ell+1)) + O(n^{-2}) \right\} \\ &= \left( \frac{1}{2n} \right)^\ell \left( 1 + \frac{\ell(\ell+2)}{2n} + O(n^{-2}) \right) \end{aligned}$$



と展開される。これらを代入すれば

$$\begin{aligned}
 F^{(\ell)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) &= \frac{1}{2^\ell} \frac{c_{n-1}}{2^{n-2}} n^{2\ell} \left\{ 1 - \ell(2\ell+3) \frac{1}{n} + O(n^{-2}) \right\} \left(\frac{1}{2n}\right)^\ell \\
 &\quad \times \left( 1 + \frac{\ell(\ell+2)}{2n} + O(n^{-2}) \right) \\
 &= \left(\frac{n}{4}\right)^\ell \frac{c_{n-1}}{2^{n-2}} \left\{ 1 - \frac{\ell(2\ell+3)}{n} + \frac{\ell(\ell+2)}{2n} + O(n^{-2}) \right\} \\
 &= \left(\frac{n}{4}\right)^\ell \frac{c_{n-1}}{2^{n-2}} \left\{ 1 - \frac{\ell(3\ell+4)}{n} + O(n^{-2}) \right\}
 \end{aligned}$$

という漸近形が導出される。因子  $n^\ell$  のため、 $n \rightarrow \infty$  での  $F^{(\ell)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right)$  の増大のスピードは微分の階数  $\ell$  が大きいほど速くなる。

$S_{2,n}$  の  $k$  次モーメント (2.2) に現れる微分を具体的に計算してみると

$$\begin{aligned}
 k=1: \quad & \frac{d}{dt} e^t F(\cdots; e^t) = e^{2t} F^{(1)}(\cdots; e^t) + e^t F^{(0)}(\cdots; e^t) \\
 k=2: \quad & \frac{d^2}{dt^2} e^t F(\cdots; e^t) = e^{3t} F^{(2)}(\cdots; e^t) + 3e^{2t} F^{(1)}(\cdots; e^t) + e^t F^{(0)}(\cdots; e^t) \\
 k=3: \quad & \frac{d^3}{dt^3} e^t F(\cdots; e^t) = e^{4t} F^{(3)}(\cdots; e^t) + 6e^{3t} F^{(2)}(\cdots; e^t) + 7e^{2t} F^{(1)}(\cdots; e^t) \\
 & \quad + e^t F^{(0)}(\cdots; e^t) \\
 k=4: \quad & \frac{d^4}{dt^4} e^t F(\cdots; e^t) = e^{5t} F^{(4)}(\cdots; e^t) + 10e^{4t} F^{(3)}(\cdots; e^t) + 25e^{3t} F^{(2)}(\cdots; e^t) \\
 & \quad + 15e^{2t} F^{(1)}(\cdots; e^t) + e^t F^{(0)}(\cdots; e^t) \\
 & \vdots
 \end{aligned}$$

となる。ただし、関数  $F$  の共通の引数  $\frac{2-n}{2}, \frac{3-n}{2}; 2$  を “...” と省略した。これらより

$$\begin{aligned}
 \frac{d^k}{dt^k} e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) &= e^{(k+1)t} F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \\
 &\quad + \frac{k(k+1)}{2} e^{kt} F^{(k-1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \\
 &\quad + ((k-2) \text{ 回以下の微分の項}) \tag{2.9}
 \end{aligned}$$

と予想できる。実際にこれが正しいことは簡単な帰納法から得られる。(2.9) で  $t=0$  とす

れば

$$\begin{aligned}
& \left. \frac{d^k}{dt^k} e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \right|_{t=0} \\
&= F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + \frac{k(k+1)}{2} F^{(k-1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\
&\quad + (k-2 \text{ 回以下の微分の項}) \\
&= F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + \frac{k(k+1)}{2} F^{(k-1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + \frac{c_{n-1}}{2^{n-2}} \times O(n^{k-2}).
\end{aligned}$$

$n \rightarrow \infty$  では微分の階数が大きい項ほど増大が大きかったので、右辺の先頭にある 2 項が支配的となる（右辺第 3 項（微分の階数が小さい項）の影響は小さい）。

以上より  $S_{2,n}$  の  $k$  次モーメントの漸近表示として

$$\begin{aligned}
E[S_{2,n}^k] &= \frac{2^{n-2}}{c_{n-1}} \left. \frac{d^k}{dt^k} e^t F\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; e^t\right) \right|_{t=0} \\
&= \frac{2^{n-2}}{c_{n-1}} F^{(k)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) + \frac{2^{n-2}}{c_{n-1}} \frac{k(k+1)}{2} F^{(k-1)}\left(\frac{2-n}{2}, \frac{3-n}{2}; 2; 1\right) \\
&\quad + O(n^{k-2}) \\
&= \left(\frac{n}{4}\right)^k \left\{ 1 - \frac{k(3k+4)}{n} + O(n^{-2}) \right\} + \left(\frac{n}{4}\right)^{k-1} \{1 + O(n^{-1})\} + O(n^{k-2}) \\
&= \left(\frac{n}{4}\right)^k \left( 1 + \frac{k^2}{2n} + O(n^{-2}) \right) \tag{2.10}
\end{aligned}$$

が得られる。

高次の枝に関する確率変数  $S_{r,n}$  のモーメントの漸近形の導出のために前節で得た漸化式 (2.8) を利用する。

$S_{3,n}$  の  $k$  次モーメントは

$$\begin{aligned}
 E[S_{3,n}^k] &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} E[S_{2,m}^k] \\
 &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} \left( \frac{m}{4} \right)^k \left( 1 + \frac{k^2}{2m} + O(m^{-2}) \right) \\
 &= \frac{1}{4^k} \frac{1}{c_{n-1}} \underbrace{\sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} m^k}_{E[S_{2,n}^k]} \\
 &\quad + \frac{1}{4^k} \frac{k^2}{2} \frac{1}{c_{n-1}} \underbrace{\sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} m^{k-1}}_{E[S_{2,n}^{k-1}]} + O(n^{k-2}) \\
 &= \frac{1}{4^k} E[S_{2,n}^k] + \frac{1}{4^k} \frac{k^2}{2} E[S_{2,n}^{k-1}] + O(n^{k-2})
 \end{aligned}$$

のように展開される。この式は、 $S_{3,n}$  の  $k$  次モーメントが  $S_{2,n}$  の  $k$  次以下のモーメントを用いて展開されることを意味する。 $S_{2,n}$  のモーメントの漸近表示 (2.10) を代入すると

$$\begin{aligned}
 E[S_{3,n}^k] &= \frac{1}{4^k} \left( \frac{n}{4} \right)^k \left( 1 + \frac{k^2}{2n} \right) + \frac{1}{4^k} \frac{k^2}{2} \left( \frac{n}{4} \right)^{k-1} + O(n^{k-2}) \\
 &= \left( \frac{n}{4^2} \right)^k \left( 1 + \frac{5k^2}{2n} \right) + O(n^{k-2})
 \end{aligned} \tag{2.11}$$

のように展開できる。

(2.11) からの類推により

$$E[S_{r,n}^k] = \left( \frac{n}{4^{r-1}} \right)^k \left( 1 + \frac{\mu_r k^2}{2n} \right) O(n^{k-2}) \tag{2.12}$$

という形を仮定する。ただし、 $\mu_r$  は  $r$  に依存した定数であり、(2.10) より  $\mu_2 = 1$ 、(2.11) より  $\mu_3 = 5$  である。 $\mu_r$  の定義より、 $S_{r+1,n}$  の  $k$  次モーメントは

$$E[S_{r+1,n}^k] = \left( \frac{n}{4^r} \right)^k \left( 1 + \frac{\mu_{r+1} k^2}{2n} \right) O(n^{k-2}) \tag{2.13}$$

で与えられる。一方、漸化式 (2.8) を用いると

$$\begin{aligned}
 E[S_{r+1,n}^k] &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} E[S_{r,m}^k] \\
 &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor n/2 \rfloor} 2^{n-2m} \binom{n-2}{n-2m} c_{m-1} \left( \frac{m}{4^{r-1}} \right)^k \left( 1 + \frac{\mu_r k^2}{2m} + O(m^{-2}) \right) \\
 &= \left( \frac{1}{4^{r-1}} \right)^k \left( E[S_{2,n}^k] + \frac{\mu_r k^2}{2} E[S_{2,n}^{k-1}] \right) + O(n^{k-2}) \\
 &= \left( \frac{n}{4^r} \right)^k \left\{ 1 + \frac{(4\mu_r + 1)k^2}{2n} \right\} + O(n^{k-2})
 \end{aligned} \tag{2.14}$$

が得られる。(2.13) と (2.14) を比較すれば、 $\mu_r$  を決定する漸化式が得られる：

$$\begin{cases} \mu_{r+1} = 4\mu_r + 1, \\ \mu_2 = 1 \quad (\text{初期条件}). \end{cases}$$

解  $\mu_r$  は

$$\mu_r = \frac{1}{3} (4^{r-1} - 1) \quad \text{for } r = 2, 3, \dots$$

これを (2.12) に代入すれば、モーメントの漸近展開の最終的な形

$$E[S_{r,n}^k] = \left( \frac{n}{4^{r-1}} \right)^k \left\{ 1 + \frac{(4^{r-1} - 1)k^2}{6n} \right\} + O(n^{k-2}) \tag{2.15}$$

に到達する。

隣接する次数のモーメントの比は漸近的に

$$\begin{aligned}
 \frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} &= \frac{\left( \frac{n}{4^{r-1}} \right)^k \left\{ 1 + \frac{(4^{r-1} - 1)k^2}{6n} \right\} + O(n^{k-2})}{\left( \frac{n}{4^r} \right)^k \left\{ 1 + \frac{(4^r - 1)k^2}{6n} \right\} + O(n^{k-2})} \\
 &= 4^k \left( 1 - \frac{4^{r-1}}{2n} k^2 \right) + O(n^{-2})
 \end{aligned} \tag{2.16}$$

と展開される。 $n \rightarrow \infty$  の極限をとれば、

$$\lim_{n \rightarrow \infty} \frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} = 4^k \tag{2.17}$$

が得られる。これは本節の冒頭で提示した式である。式 (2.17) より、隣接する次数に関する  $k$  次モーメントの比の極限值  $4^k$  は  $r$  によらない。

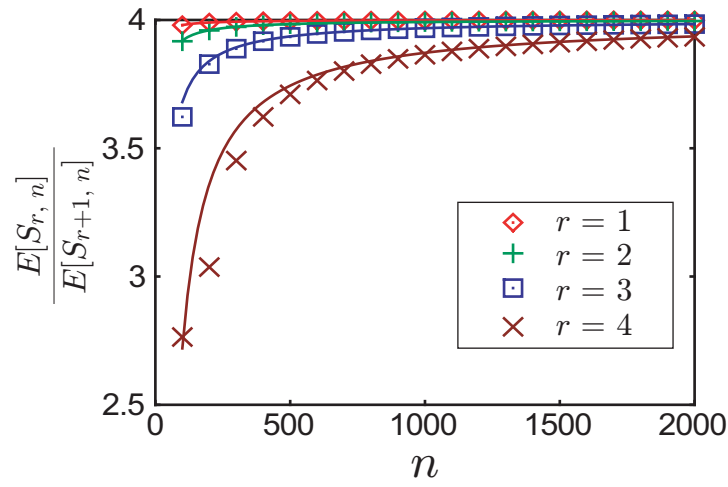


図 2.6 分岐比の漸近形  $\frac{E[S_{r,n}]}{E[S_{r+1,n}]} \simeq 4 - \frac{4^r}{2n}$  と漸化式の数値計算結果の比較。各点が漸化式から得られる分岐比、曲線が関数  $4 - \frac{4^r}{2n}$ 。

式 (2.16) で  $k = 1$  とすれば

$$\frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} = 4 - \frac{4^r}{2n} + O(n^{-2})$$

となる。この式で  $O(n^{-2})$  の項を無視した漸近曲線と漸化式 (2.8) の数値計算結果の比較を図 2.6 に示す。式 (2.17) において  $k = 1$  の場合には

$$\lim_{n \rightarrow \infty} \frac{E[S_{r,n}]}{E[S_{r+1,n}]} = 4 \quad (2.18)$$

が得られる (ランダムモデルでは分岐比が漸近的に 4 である)。すなわち、ランダムモデルはサイズが大きい場合に位相的自己相似性をもつ。3.4 節ではランダムモデルがさらに強い位相的自己相似性をもつことを述べる。

## 第3章 期待値の漸近展開公式およびランダムモデルの位相的自己相似性

高次の枝に関する確率変数  $S_{r,n}$  のモーメントの漸近的な性質を調べるうえで、漸化式 (2.8) が重要であった。しかし、前節で示した (2.8) の導出は (2.6) 等を経由した非常に複雑なものであった。本章では二分木に対するある種の粗視化変換を導入し、その性質を利用して (2.8) よりも一般性のある漸化式を導く。この漸化式から一般的な確率変数の期待値の漸近展開の公式を導出する。さらに、ランダムモデルが強い位相的自己相似性をもつことを述べ、いくつかの計算例を示す。

### 3.1 一般的な確率変数について

$\Omega_n$  に属する二分木の  $r$  次の枝の本数を与える確率変数  $S_{r,n}$  は Horton-Strahler 解析において基本的かつきわめて重要である。本章では一般化した確率変数として  $f(S_{r,n})$  と表されるものを考える。 $S_{r,n}$  は非負整数値しかとらないので関数  $f$  の定義域も  $\mathbb{N} \cup \{0\}$  であれば充分だが、解析的な扱いやすさのため定義域および値域がともに実数全体となるものを考える（さらに、3.3 節以降では無限遠点まわりの Laurent 展開が可能である  $f$  のみを扱う）。確率変数  $S_{r,n}$  自体が  $S_{r,n} : \Omega_n \rightarrow \mathbb{N} \cup \{0\}$  という関数なので、 $f(S_{r,n})$  は 2 つの関数  $S_{r,n}$  と  $f$  の合成関数であり、 $f(S_{r,n}) : \Omega_n \rightarrow \mathbb{R}$  という実数値の確率変数である。すなわち、二分木  $T \in \Omega_n$  に対して 1 つの実数  $f(S_{r,n}(T))$  が定まる。以下の節では一般的な確率変数  $f(S_{r,n})$  の期待値  $E[f(S_{r,n})]$  の漸近表示を導出する。

例えば、 $f(x) = x^k$  の場合、 $E[f(S_{r,n})]$  は  $S_{r,n}$  の  $k$  次モーメントである。また、 $f(x) = (x - E[x])^2$  の場合、 $E[f(S_{r,n})]$  は  $S_{r,n}$  の分散である。

### 3.2 二分木の粗視化

二分木  $T \in \Omega_n$  に対して、次の 2 つのステップ

- (i)  $T$  の頂点の中で次数 1 のものを全て除去する（すなわち、 $T$  の全ての葉を除去する）。

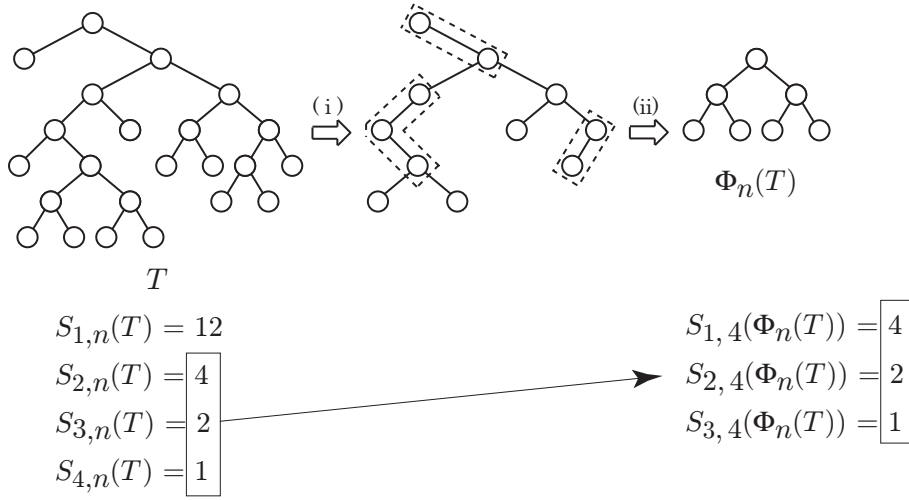


図 3.1 変換  $\Phi_n$  の例 ( $n = 12$ )。 (i)  $T$  から全ての葉を除去する。 (ii) 枝の縮約。破線で囲まれた部分を 1 つの頂点に置き換える。 $T$  の  $r$  次の枝の数  $S_{r,n}(T)$  と  $\Phi_n(T)$  の  $r - 1$  次の枝の数  $S_{r-1,n}(\Phi_n(T))$  は等しい (式 (3.1) を参照)。

(ii) 操作 (i) によって子を 1 つしかもたない頂点が生じた場合、その頂点と子を結ぶ辺を縮約する。

によりつくられた新たな二分木を  $\Phi_n(T)$  と書くことにする。このとき、 $\Phi_n$  はサイズ  $n$  の二分木から別の二分木への変換となっている (図 3.1 に具体例を示す)。 $\Phi_n(T)$  のサイズは高々  $\lfloor n/2 \rfloor$  個である。したがって定義域および値域を明確にすると、 $\Phi_n$  は

$$\Phi_n : \Omega_n \rightarrow \bigcup_{m=1}^{\lfloor n/2 \rfloor} \Omega_m$$

という写像である。 $\Phi_n$  は全射だが、一対一ではない。

2 次の枝の個数に応じて  $\Omega_n$  を分割する。 $\Omega_n^m = \{T \in \Omega_n \mid S_{2,n}(T) = m\}$  とおく (つまり、 $\Omega_n^m$  はサイズ  $n$  の二分木のうち 2 次の枝が  $m$  個であるものの全体である)。これはまた、 $\Omega_n^m = \Phi_n^{-1}(\Omega_m)$  とも表せる。 $\Omega_n$  に属する二分木の 2 次の枝の個数は最大  $\lfloor n/2 \rfloor$  個なので

$$\Omega_n = \bigcup_{m=1}^{\lfloor \frac{n}{2} \rfloor} \Omega_n^m, \quad \Omega_n^m \cap \Omega_n^{m'} = \emptyset \quad (m \neq m')$$

である。すなわち  $\{\Omega_n^m\}_{m=1}^{\lfloor n/2 \rfloor}$  は  $\Omega_n$  の直和分解になっている。これらを図的に表現したのが図 3.2 である。

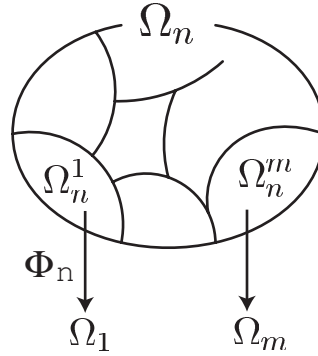


図 3.2 集合  $\Omega_n^m$  の概念図。写像  $\Phi_n$  の作用により得られる二分木のサイズ  $m$  に応じて  $\Omega_n$  の分割  $\{\Omega_n^m\}_m$  が与えられる。

$\Phi_n$  は次数 1 の枝を除去する変換であるため、二分木  $T \in \Omega_n$  の 2 次の枝が二分木  $\Phi_n(T)$  の 1 次の枝に対応する。この対応関係は高次の枝でも成立し、二分木  $T \in \Omega_n^m$  に対して

$$S_{r,n}(T) = S_{r-1,m}(\Phi_n(T)) \quad (r \geq 2) \quad (3.1)$$

が成立する。つまり、 $r$  次の確率変数  $S_{r,n}$  と  $(r-1)$  次の確率変数  $S_{r-1,m}$  が  $\Phi_n$  を介して結びついている。例えば、図 3.1 に示した二分木の場合、 $n = 12$ ,  $m = 4$  であり、

$$S_{2,n}(T) = S_{1,m}(\Phi_n(T)) = 4,$$

$$S_{3,n}(T) = S_{2,m}(\Phi_n(T)) = 2,$$

$$S_{4,n}(T) = S_{3,m}(\Phi_n(T)) = 1$$

となっている。

$\Phi_n$  の定義域を  $\Omega_n^m$  に制限した写像  $\Phi_n|_{\Omega_n^m} : \Omega_n^m \rightarrow \Omega_m$  を考える。この写像も一対一ではなく、 $\Omega_n^m$  に属するいくつかの二分木が  $\Phi_n$  によって同一の二分木に変換される (図 3.3)。そこで、 $\tau \in \Omega_m$  として  $\Phi_n(T) = \tau$  となるような  $T \in \Omega_n$  の総数を

$$\mu_n^m(\tau) := \#\{T \in \Omega_n | \Phi_n(T) = \tau\} (\equiv \#\Phi_n^{-1}(\tau))$$

と表す。変換  $\Phi_n$  の性質は多重度  $\mu_n^m(\tau)$  に集約されている。 $\mu_n^m(\tau)$  を計算するために  $\Phi_n$  の逆写像を具体的に数え上げる。大雑把にいて  $\Phi_n$  は二分木の  $n$  個の葉を除去する操作だったので、逆像  $\Phi_n^{-1}(\tau)$  は二分木  $\tau (\in \Omega_m)$  に  $n$  個の葉をくっつけることで構成できる。具体的には以下の 3 つのプロセスからなる。

- (1)  $\tau$  の全ての葉に 2 つずつ新たな葉を取り付ける。この操作により  $\tau$  の全ての葉が 2 次の枝になる。



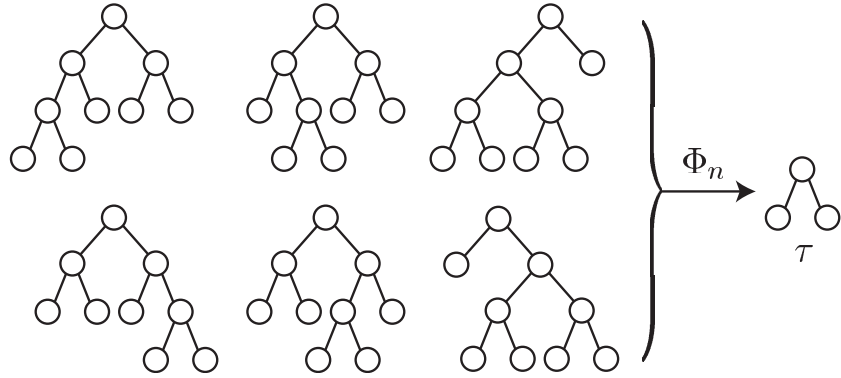


図 3.3 変換  $\Phi_n$  による二分木の縮退の例 ( $n = 5$ )。  $\Omega_5$  に属する 6 つの二分木が  $\Phi_5$  によって同一の二分木  $\tau (\in \Omega_2)$  に変換される。この場合の多重度は  $\mu_5^2(\tau) = 6$  である。

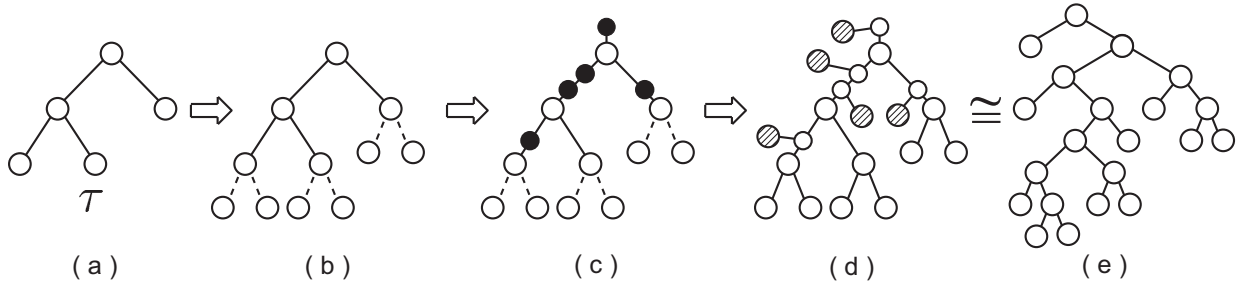


図 3.4 逆像  $\Phi_n^{-1}(\tau)$  の構成の例 ( $m = 3, n = 11$ )。 (a) 二分木  $\tau \in \Omega_m$ . (b)  $\tau$  の葉に新たな葉を 2 つずつ付ける. (c)  $n - 2m (= 5)$  個の残りの葉をどの辺に取り付けるか決める. (d) それぞれの葉を左右どちらからぶら下げるかを選ぶ. (e) 結果として得られる二分木

(2) (1) で  $2m$  個の葉を取り付けたので、残り  $n - 2m$  個の葉を取り付ける必要がある。これらをどの辺に付けるかを選ぶ (根の直上に付けることも可能である)。

(3)  $n - 2m$  個の葉を左右どちらからぶら下げるかを選ぶ。

逆像の構成例 ( $m = 3, n = 11$ ) が図 3.4 である。  $\Omega_m$  の二分木の辺の本数は  $2m - 2$  である。上の手順 (2) では  $n - 2m$  個の葉を (根の上も含めて)  $2m - 1$  個の辺に配分するので、 $\binom{n-2}{n-2m}$  とおりのやり方がある。また、(3) では  $n - 2m$  個の葉のぶら下げる方向 (左か右) を指定できるので、可能なぶら下げ方は  $2^{n-2m}$  だけある。これら 2 つの因子は独立なので、多重度は

$$\mu_n^m(\tau) = \binom{n-2}{n-2m} 2^{n-2m} \quad (3.2)$$

として与えられる。式 (3.2) の右辺は  $n$  および  $m$  のみしか現れていない。これは、多重度  $\mu_n^m(\tau)$  が二分木  $\tau \in \Omega_m$  によらず一定であることを意味する。このことから、以下では

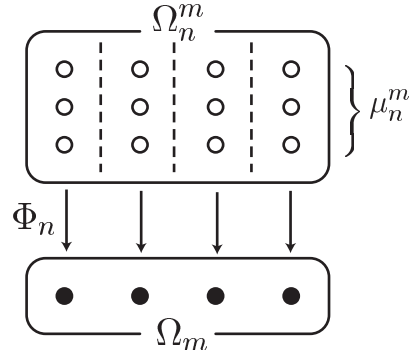


図 3.5  $\Omega_n^m$ ,  $\Omega_m$ ,  $\mu_n^m$  の関係を表す模式図。白丸、黒丸はそれぞれ  $\Omega_n^m$  および  $\Omega_m$  の要素を表す。縦に並んだ白丸は  $\Phi_n$  によって同一の黒丸にうつる。多重度  $\mu_n^m(\tau)$  が  $\tau \in \Omega_m$  によらず一定であることから、 $\Omega_n^m$  の元を長方形に配置することができる。

$\mu_n^m(\tau)$  を単に  $\mu_n^m$  と表記する。ちなみに、 $\mu_n^m$  は前章(式(2.1))で“ぶら下げ”を用いて導出した  $\mathcal{N}(n, m)$  と同一のものである。全ての  $\tau \in \Omega_m$  に対して逆像  $\Phi_n^{-1}(\tau)$  が同じ個数の元を含むので、図 3.5 に示すように  $\Omega_n^m$  の元を長方形に並べることができる。以上より、 $\Omega_n^m$ ,  $\Omega_m$  の要素数および  $\mu_n^m$  に対して

$$\mu_n^m = \frac{\#\Omega_n^m}{\#\Omega_m}$$

という関係が成立する。 $\Omega_n^m$  の要素数は

$$\#\Omega_n^m = \mu_n^m \cdot \#\Omega_m = \binom{n-2}{n-2m} 2^{n-2m} \frac{(2m-2)!}{m!(m-1)!} = \frac{(n-2)! 2^{n-2m}}{(n-2m)! m! (m-1)!}$$

と明示的に表される。

確率変数  $f(S_{r,n})$  の期待値の表式を変形していくと

$$\begin{aligned}
E[f(S_{r,n})] &= \frac{1}{c_{n-1}} \sum_{T \in \Omega_n} f(S_{r,n}(T)) \\
&= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{T \in \Omega_n^m} f(S_{r,n}(T)) \\
&= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{T \in \Omega_n^m} f(S_{r-1,m}(\Phi_n(T))) \\
&= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m \sum_{\tau \in \Omega_m} f(S_{r-1,m}(\tau)) \\
&= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m c_{m-1} \frac{1}{c_{m-1}} \sum_{\tau \in \Omega_m} f(S_{r-1,m}(\tau)) \\
&= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m c_{m-1} E[f(S_{r-1,m})] \\
&= \frac{n!(n-1)!(n-2)!}{(2n-2)!} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \frac{2^{n-2m}}{(n-2m)!m!(m-1)!} E[f(S_{r-1,m})] \quad (3.3)
\end{aligned}$$

という形が得られる。右辺には  $r-1$  次の量  $E[f(S_{r-1,m})]$  があるため、この式は次数  $r$  に関する漸化式である。 $f(x) = x^k$  の場合が前章の漸化式 (2.8) である。この式は変換  $\Phi_n$  に関連した諸性質 (3.1), (3.2) 等を効果的に利用することで (2.8) に比べて見通しよく導出されていることに注意。

### 3.3 期待値 $E[f(S_{r,n})]$ の漸近挙動

前節で得た漸化式 (3.3) を用いて、一般的な確率変数  $f(S_{r,n})$  の期待値の  $n \rightarrow \infty$  でふるまいを調べる。1 次の枝は二分木の葉にほかならないので  $S_{1,n}(T) \equiv n$  である。すなわち、確率変数  $f(S_{1,n})$  は  $\Omega_n$  上で定数値  $f(n)$  をとる。よって期待値は  $E[f(S_{1,n})] = E[f(n)] = f(n)$  であり、これが漸化式 (3.3) の初期条件に相当する。関数  $f(n)$  は次の形に展開できると仮定する:

$$E[f(S_{1,n})] = f(n) = a_1 n^k + b_1 n^{k-1} + O(n^{k-2}). \quad (3.4)$$

これは、関数  $f$  を複素関数 ( $f: \mathbb{C} \rightarrow \mathbb{C}$ ) に拡張した場合の無限遠点での Laurent 展開である。多項式関数や有理関数など、関数  $f$  が無限遠に真性特異点をもたなければ展開 (3.4)

が可能である。3.4 節に示す例のように、實際上重要なのは  $f$  が多項式または有理関数である場合である。

高次の確率変数  $f(S_{r,n})$  についても (3.4) と同様に展開できると仮定する:

$$E[f(S_{r,n})] = a_r n^k + b_r n^{k-1} + O(n^{k-2}). \quad (3.5)$$

ただし、係数  $a_r$  および  $b_r$  は  $n$  によらない。 $a_r, b_r$  を初期条件に現れる  $a_1, b_1$  および  $k$  を用いて表したものが漸近展開の公式である。

前章の式 (2.10) より、確率変数  $S_{2,n}$  の  $p$  次モーメントは

$$E[S_{2,n}^p] = \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m c_{m-1} m^p = \left(\frac{n}{4}\right)^p \left(1 + \frac{p^2}{2n}\right) + O(n^{p-2})$$

と書かれていた。これを用いて漸化式 (3.3) を変形していくと

$$\begin{aligned} E[f(S_{r,n})] &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m E[f(S_{r-1,m})] \\ &= \frac{1}{c_{n-1}} \sum_{m=1}^{\lfloor \frac{n}{2} \rfloor} \mu_n^m c_{m-1} \{a_{r-1} m^k + b_{r-1} m^{k-1} + O(m^{k-2})\} \\ &= a_{r-1} E[S_{2,n}^k] + b_{r-1} E[S_{2,n}^{k-1}] \\ &= a_{r-1} \left(\frac{n}{4}\right)^k \left(1 + \frac{k^2}{2n}\right) + b_{r-1} \left(\frac{n}{4}\right)^{k-1} + O(n^{k-2}) \\ &= \frac{a_{r-1}}{4^k} n^k + \left(\frac{b_{r-1}}{4^{k-1}} + \frac{k^2 a_{r-1}}{2 \cdot 4^k}\right) n^{k-1} + O(n^{k-2}) \end{aligned} \quad (3.6)$$

のようになる（最右辺では  $(r-1)$  次の量  $a_{r-1}, b_{r-1}$  のみが残ることに注意）。

式 (3.5) および (3.6) を比較する。 $O(n^k)$  の項より、数列  $\{a_r\}_{r \in \mathbb{N}}$  の漸化式

$$a_r = \frac{a_{r-1}}{4^k}$$

が得られる。これは簡単に解くことができ、一般解は

$$a_r = \left(\frac{1}{4^k}\right)^{r-1} a_1 \quad (3.7)$$

である。

次に  $O(n^{k-1})$  の項を比較すると漸化式

$$b_r = \frac{b_{r-1}}{4^{k-1}} + \frac{k^2 a_{r-1}}{2 \cdot 4^k} = \frac{b_{r-1}}{4^{k-1}} + \frac{k^2 a_1}{2} \frac{1}{4^{k(r-1)}} \quad (3.8)$$

が得られる (2 番目の等号では (3.7) を用いた)。ここで、

$$U = \frac{1}{4^{k-1}}, \quad V = \frac{k^2 a_1}{2}, \quad W = \frac{1}{4^k} \quad (3.9)$$

とおくと (3.8) はシンプルな形

$$b_r = Ub_{r-1} + VW^{r-1}$$

に帰着する。初期値  $b_1$  から順に計算していくと

$$\begin{aligned} b_2 &= Ub_1 + VW \\ b_3 &= Ub_2 + VW^2 = U^2 b_1 + V(UW + W^2) \\ b_4 &= Ub_3 + VW^3 = U^3 b_1 + V(U^2 W + UW^2 + W^3) \\ b_5 &= Ub_4 + VW^4 = U^4 b_1 + V(U^3 W + U^2 W^2 + UW^3 + W^4) \\ &\vdots \end{aligned}$$

これより  $b_r$  は

$$\begin{aligned} b_r &= U^{r-1} b_1 + V \sum_{s=0}^{r-2} U^s W^{r-1-s} = U^{r-1} b_1 + VW^{r-1} \sum_{s=0}^{r-2} \left( \frac{U}{W} \right)^s \\ &= U^{r-1} b_1 + VW^{r-1} \frac{\left( \frac{U}{W} \right)^{r-1} - 1}{\frac{U}{W} - 1} \end{aligned}$$

と表せる。ただし、(3.9) より  $U/W = 4$  となることを見越して  $U/W$  という因子を残してある。 $U, V, W$  を元の形に戻せば漸化式 (3.8) の一般解

$$b_r = \frac{b_1}{4^{(k-1)(r-1)}} + \frac{k^2 a_1}{2} \frac{1}{4^{k(r-1)}} \frac{4^{r-1} - 1}{3} = \left( \frac{1}{4^{r-1}} \right)^k \left\{ 4^{r-1} b_1 + \frac{4^{r-1} - 1}{6} k^2 a_1 \right\} \quad (3.10)$$

が得られる。

数列  $\{a_r\}_r$  および  $\{b_r\}_r$  の解 (3.7) および (3.10) を (3.5) に代入すれば

$$E[f(S_{r,n})] = \left( \frac{n}{4^{r-1}} \right)^k \left\{ a_1 + \frac{1}{n} \left( 4^{r-1} b_1 + \frac{4^{r-1} - 1}{6} k^2 a_1 \right) \right\} + O(n^{k-2}) \quad (3.11)$$

という漸近展開が得られる。 $r = 1$  とすれば初期条件 (3.4) と一致することにも注意せよ。

関数  $f$  が多変数の場合も同様の結果が成り立つ。例えば、2 変数関数  $f(x_1, x_2)$  から確率変数  $f(S_{r,n}, S_{r+1,n})$  をつくったとする。 $r = 1$  の場合に

$$E[f(S_{1,n}, S_{2,n})] = E[f(n, S_{2,n})] = a_1 n^k + b_1 n^{k-1} + O(n^{k-2})$$

という展開をもつならば、高次の確率変数に対して (3.11) と同じ形の漸近展開

$$E[f(S_{r,n}, S_{r+1,n})] = \left(\frac{n}{4^{r-1}}\right)^k \left\{ a_1 + \frac{1}{n} \left( 4^{r-1} b_1 + \frac{4^{r-1}-1}{6} k^2 a_1 \right) \right\} + O(n^{k-2}) \quad (3.12)$$

が得られる。

### 3.4 位相的自己相似性とその一般化

ここで、本章で導入した“二分木の葉をむしり取る”変換  $\Phi_n$  の意味を考えてみる。植物や河川網など実際の分岐パターンの多くは次数が小さい（つまり、末端に近い）枝ほど細く短い。変換  $\Phi_n$  は二分木の細かいスケールの構造を無視すること、すなわち粗視化に対応する。ある二分木  $T \in \Omega_n$  およびその粗視化  $\Phi_n(T)$  について、分岐の度合いが似ているとする。つまり、 $T$  と  $\Phi_n(T)$  は  $r$  次の分岐比が同じ値だとする（分岐比が枝分かれの複雑さの度合いを表していた）。式 (3.1) より、 $\Phi_n(T)$  の  $r$  次の分岐比は  $T$  の  $(r+1)$  次の分岐比と等しい。これより、“二分木の自己相似性”として分岐比が次数によらない定数となることとみなすのは妥当である。このように、粗視化  $\Phi_n$  を通して眺めると位相的自己相似性の概念をより明確にとらえることができる。

ランダムモデルにおいては分岐比が次数によらず漸近的に 4 であるという結果 (2.18) が得られた。また、 $k$  次モーメントの比も次数  $r$  によらないという“相似性” (2.17) が成り立っていた。

通常分岐比を一般化し、関数  $f$  を介した分岐比

$$R_{r,n}^f = \frac{E[f(S_{r,n})]}{E[f(S_{r+1,n})]}$$

の漸近挙動を調べる。恒等関数  $f(x) = x$  の場合に  $R_{r,n}^f$  は従来の分岐比と一致する。

漸近展開 (3.11) より、 $f$  を介した分岐比は

$$\begin{aligned} R_{r,n}^f &= \frac{E[f(S_{r,n})]}{E[f(S_{r+1,n})]} \\ &= \frac{\left(\frac{n}{4^{r-1}}\right)^k \left\{ a_1 + \frac{1}{n} \left( 4^{r-1} b_1 + \frac{4^{r-1}-1}{6} k^2 a_1 \right) \right\} + O(n^{k-2})}{\left(\frac{n}{4^r}\right)^k \left\{ a_1 + \frac{1}{n} \left( 4^r b_1 + \frac{4^r-1}{6} k^2 a_1 \right) \right\} + O(n^{k-2})} \\ &= 4^k - \frac{4^{k+r-1}(6b_1 + a_1 k^2)}{2a_1 n} + O(n^{-2}) \end{aligned} \quad (3.13)$$

と展開される。特に、 $\lim_{n \rightarrow \infty} R_{r,n}^f = 4^k$  であり、極限值は  $f$  の先頭次数  $k$  のみにしか依存しない。この極限值が次数  $r$  によらず一定であるということは位相的自己相似性をより一

般化した主張である ( $f(x)$  が恒等関数のとき従来の位相的自己相似性に帰着する)。したがってサイズが充分大きなランダムモデルは、より強い位相的自己相似性をもつことが示された。 $f$  が多変数関数の場合にも展開公式 (3.12) を用いれば (3.13) と同様の結果が得られる。

以下、関連した計算結果を示す。

- (a)  $f(x) = x^k$  ( $k \in \mathbb{N}$ ) の場合を考える。 $E[f(S_{r,n})] = E[S_{r,n}^k]$  は  $S_{r,n}$  の  $k$  次モーメントである。すでに  $f$  は (3.4) の形になっており、ただちに  $a_1 = 1$  および  $b_1 = 0$  が分かる。漸近展開 (3.11) は

$$E[S_{r,n}^k] = \left(\frac{n}{4^{r-1}}\right)^k \left(1 + \frac{4^{r-1} - 1}{6n} k^2\right) + O(n^{k-2}) \quad (3.14)$$

である。前章の結果 (2.15) が公式 (3.11) から再導出された。また、(3.13) に代入すれば

$$\frac{E[S_{r,n}^k]}{E[S_{r+1,n}^k]} = 4^k - \frac{4^{k+r-1} k^2}{2n} + O(n^{-2})$$

が得られる。これも (2.16) の再導出である。

- (b)  $S_{r,n}$  の分散  $\text{var}(S_{r,n})$  の漸近的な性質を考える。式 (3.11) を用いると

$$\begin{aligned} \text{var}(S_{r,n}) &= E[S_{r,n}^2] - E[S_{r,n}]^2 \\ &= \left(\frac{n}{4^{r-1}}\right)^2 \left(1 + \frac{4^{r-1} - 1}{6n} 4\right) + O(n^0) - \left[\frac{n}{4^{r-1}} \left(1 + \frac{4^{r-1} - 1}{6n}\right) + O(n^{-1})\right]^2 \\ &= \frac{1}{3} \frac{4^{r-1} - 1}{16^{r-1}} n + O(n^0) \end{aligned}$$

が得られる。 $O(n^2)$  の項が打ち消されていることに注意。ゆえに、

$$\frac{\text{var}(S_{r,n})}{n} \longrightarrow \frac{1}{3} \frac{4^{r-1} - 1}{16^{r-1}} \quad \text{as } n \rightarrow \infty \quad (3.15)$$

という極限值が得られる。これは Moon [50] による結果 (1.4) を再現する。この結果は第5章で述べる中心極限定理の数値計算結果の妥当性の根拠の1つとして用いる。

- (c)  $f$  が多変数である場合の例として  $f(S_{r,n}, S_{r+1,n}) = S_{r+1}/S_{r,n}$  を取り上げる。ちなみに、この確率変数を逆分岐比とよぶ (第5章)。  $r = 1$  の場合、 $S_{2,n}$  の期待値 (2.3a) を利用すると

$$E\left[\frac{S_{2,n}}{S_{1,n}}\right] = E\left[\frac{S_{2,n}}{n}\right] = \frac{1}{n} E[S_{2,n}] = \frac{n-1}{2(2n-3)} = \frac{1}{4} + \frac{1}{8n} + O(n^{-2})$$

これより、 $a_1 = \frac{1}{4}$ ,  $b_1 = \frac{1}{8}$ ,  $k = 0$  となる。(3.12) に代入すれば

$$E \left[ \frac{S_{r+1,n}}{S_{r,n}} \right] = \frac{1}{4} + \frac{4^{r-2}}{2n} + O(n^{-2}) \quad (3.16)$$

が得られる。一方、(3.14) より

$$\frac{E[S_{r+1,n}]}{E[S_{r,n}]} = \frac{1}{4 - \frac{4^r}{2n} + O(n^{-2})} = \frac{1}{4} + \frac{4^{r-2}}{2n} + O(n^{-2}) \quad (3.17)$$

が得られる。 $r \geq 2$  の場合、(3.16) と (3.17) の左辺は等しくないのだが、 $n^{-1}$  までの展開では完全に一致している。



## 第4章 ランダムモデルの数値計算法

本章では一般の確率変数  $f(S_{r,n})$  の値を数値的に計算する手法およびその数学的背景を述べる。

### 4.1 数値計算の必要性

前章の漸近展開公式 (3.11) を用いるとかなり広いクラスの  $f$  に対して  $E[f(S_{r,n})]$  の漸近表示を計算することができる。期待値は確率変数の性質を調べるための重要な量の1つではあるが、確率変数は期待値のみで特徴付けられるわけではない。例えば、 $f(S_{r,n})$  の分布関数は期待値より大きな情報ををもっているが、公式 (3.11) では計算できない。解析的な計算が不可能である（あるいは見通しが立たない）場合、数値計算が有力な手段となる。

数値計算法として最も単純なのが、 $\Omega_n$  に属する全ての二分木をくまなく走査していく方法（全探索）である。 $\Omega_n$  の要素の総数は Catalan 数  $c_{n-1}$  で表されていた（1.5.2 節）。例えば、サイズ  $n = 100$  の位相的に異なる二分木の総数は  $\#\Omega_{100} \simeq 2.275 \times 10^{56}$  という膨大な数になってしまい、現実的な時間では全ての二分木にわたって調べ上げるのは不可能である（1秒で二分木1億個を探索できるとしても  $\Omega_{100}$  の全ての二分木を調べるのに  $7.21 \times 10^{40}$  年を要する）。なお、Stirling の公式

$$n! \simeq \sqrt{2\pi n} \frac{n^n}{e^n}$$

を用いると、 $n$  が充分大きい場合の近似式

$$\#\Omega_n = c_{n-1} = \frac{1}{n} \frac{(2n-2)!}{(n-1)!(n-1)!} \simeq \frac{4^n}{n^{3/2}\sqrt{\pi}}$$

が得られる。これより、 $\#\Omega_n$  は  $n$  に対してほぼ指数的に増大することがわかる（典型的な組合せ爆発である）。

以下、本章ではランダムモデルに対するモンテカルロ的なアルゴリズムについて説明する。二分木を一樣な確率で生成する代わりに、それと等価なビット列（Dyck 列）を等確率で生成するという方針をとる。

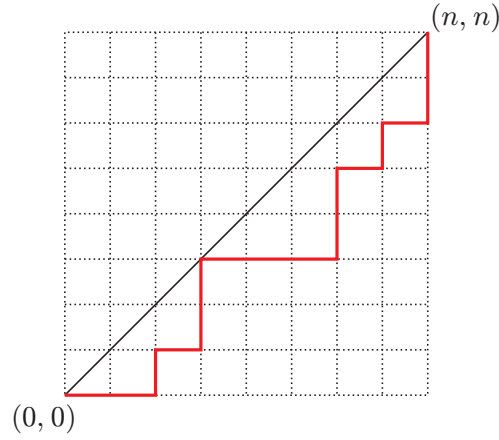


図 4.1 長さ 16 の Dyck 経路の例。(0,0) から  $(n,n)$  に引かれた対角線を上回ることはない。

## 4.2 Dyck 経路と Catalan 数

2 次元の正方格子  $\mathbb{Z}^2$  上の隣接する格子点を順に並べた点列を格子路と呼ぶ。 $\ell + 1$  個の点からなる格子路の長さは  $\ell$  であると約束する。長さ  $2n$  の格子路  $(s_0, s_1, \dots, s_{2n})$  で、2 条件

- (1)  $s_0 = (0, 0)$ ,  $s_{2n} = (n, n)$
- (2) 各格子点の座標  $s_i = (x_i, y_i)$  は  $x_i \geq y_i$  を満たす

をとともに満足するものを Dyck 経路と呼ぶ。言い方をかえれば、 $(0, 0)$  から出発し、右上がりの対角線より上にはこないように格子上を右または上へのみ遷移して  $(n, n)$  へ到達する経路である。図 4.1 も参照せよ。以下、最短の格子路のことを単に“経路”や“格子路”と呼ぶことにする。

長さ  $2n$  の Dyck 経路の総数を計算しておく。ここでは、よく知られた反射原理を利用する方法 [67, 68] を紹介する。まず、 $(0, 0)$  から最短で  $(n, n)$  に到達する格子路の総数は  $\binom{2n}{n}$  である。次に、 $(0, 0)$  と  $(n, n)$  を結ぶ経路のうち、Dyck 経路でないものの総数を数える。定義より、Dyck 経路でない格子路は少なくとも 1 度は対角線のすぐ上の格子点を通る。このような点のうち最後に通った格子点に注目し、その点から後の経路を  $y = x + 1$  の直線に関して反転させて新たな経路をつくる（“最後の格子点”は必ず直線  $y = x + 1$  上に存在する）。この経路は  $(0, 0)$  と  $(n - 1, n + 1)$  を結ぶ経路となる（図 4.2）。重要なのが  $(0, 0)$  と  $(n, n)$  を結ぶ Dyck 経路でない格子路が  $(0, 0)$  と  $(n - 1, n + 1)$  を結ぶ全ての格子路と 1 対 1 に対応するということである。したがって、Dyck 経路でない格子路の総数が  $(0, 0)$  と  $(n - 1, n + 1)$  を結ぶ経路の総数と等しく  $\binom{2n}{n - 1}$  であることが分かった。以上

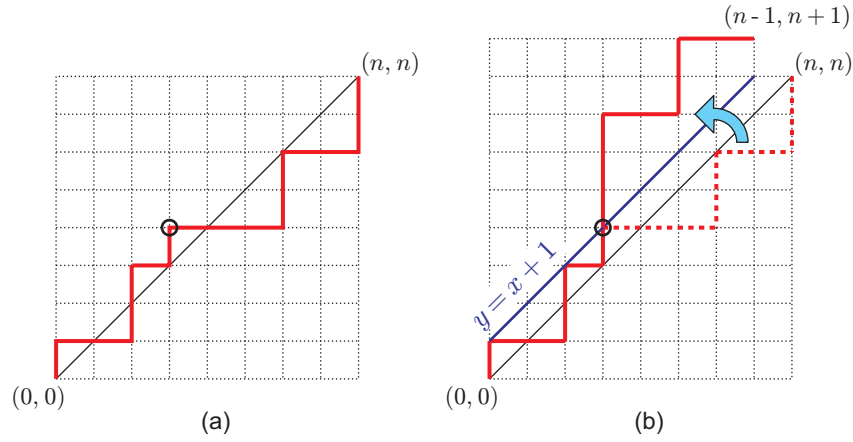


図 4.2 (a) Dyck 経路でない格子路。対角線より上にある最後の格子点を丸で囲んである。(b) 丸をつけた点より後ろの経路を直線  $y = x + 1$  に関して反転させると  $(0, 0)$  から  $(n - 1, n + 1)$  への経路が得られる。

より、長さ  $2n$  の Dyck 経路の総数は

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1} \binom{2n}{n} = c_n$$

のように  $n$  番目の Catalan 数で与えられる。サイズ  $n$  の位相的に異なる二分木の総数は  $c_{n-1}$  であったので、サイズ  $n$  の二分木と長さ  $2(n-1)$  の Dyck 経路が 1 対 1 の対応をもつ。以下、特に断らない限り Dyck 経路といえば長さ  $2(n-1)$  のものを指すとする。

さて、平面格子点上の右への遷移を ‘1’、上への遷移を ‘0’ と表せば各々の Dyck 経路が 1 と 0 で構成された文字列 (ビット列) に対応する。この操作によって得られるビット列を本論文では Dyck 列と呼ぶことにする (正確には “アルファベット  $\{0, 1\}$  上の Dyck 語” [46] という)。Dyck 経路の定義を焼き直せば、Dyck 列は明らかに次の 2 つの性質をもつ: (1) 1 つの Dyck 列に含まれる ‘0’ の個数および ‘1’ の個数はともに  $n-1$  である。(2) Dyck 列の先頭から ‘0’、‘1’ の個数をそれぞれ数えていったとき、‘0’ の個数はつねに ‘1’ の個数以下である。Dyck 列は文字列であるため、コンピュータ上では二分木よりも扱いやすい。そこで、Dyck 列 (または Dyck 経路) と二分木を 1 対 1 に対応づける方法を次節で述べる。

### 4.3 二分木と Dyck 列の対応

河川網は流路の合流が繰り返され河口へ到達する。河川網の位相構造をモデル化した二分木にも同様の ‘合流’ の視点を与えることができる。二分木において 2 つの頂点を括弧 ‘( )’ で

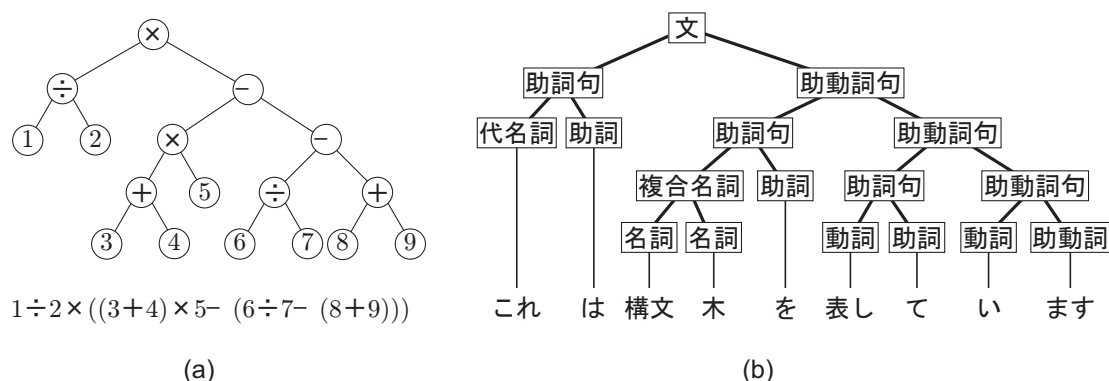


図 4.3 式木 (a) によって演算の順序が、構文木 (b) によって単語の係り受け関係が図式化される。

くくることで頂点の合流を表すことにする。すると、二分木  $T \in \Omega_n$  の構造は葉  $v_1, \dots, v_n$  および合流を表す  $()$  を用いて表現できる。例えば、 $((v_1 v_2) v_3)$  は「頂点  $v_1$  と  $v_2$  が合流した頂点と  $v_3$  が合流する」という関係を表す。二分木のもつ「合流」の性質は二項演算の計算順序を表す式木 [69, 70] や単語の係り受けの関係を図式化した構文木 [71] などに応用されている (図 4.3)。

以下の操作により二分木  $T \in \Omega_n$  から Dyck 列を得ることができる：

- (1)  $T$  を  $v_1, \dots, v_n$  および  $()$  で表す。
- (2) できた記号列から  $v_1$  および  $()$  を除去し、 $v_2, \dots, v_n$  を  $'1'$  に、 $)$  を  $'0'$  に置き換える。

こうして得られるビット列が Dyck 列となっていることは次のように分かる。まず、 $'1'$  および  $'0'$  の個数がそれぞれ  $n - 1$  であるのは明らかである。そして、演算  $()$  は引数を 2 つ必要とするので、(1) で得た記号列において最初の  $)$  は  $v_2$  よりも右側にある。同様の理由から  $m$  番目の  $)$  は  $v_{m+1}$  よりも右側にある。つまり、対応するビット列の先頭から数えていった  $'0'$  の個数は  $'1'$  の個数を上回ることがない。ゆえに、このビット列は Dyck 列である。二分木からビット列を得るプロセスの例を図 4.4 に示す。

二分木と Dyck 経路を対応させる上記の方法は二項演算子の逆ポーランド記法 [72] を利用したものと本質的に等価である。逆ポーランド記法は演算子を被演算子の後ろに記述する方法で後置記法とも呼ばれる。二分木の場合、頂点の合流が二項演算に相当し、上の方法で  $()$  を除去し  $)$  を残したことが演算子の後置に対応する。

二分木の各頂点の次数の計算を、対応する Dyck 列を用いて行なうことができる。

- (1) Dyck 列の先頭に  $'1'$  を 1 つ加える (二分木から Dyck 列をつくる際に除去した葉  $v_1$  を戻す)。

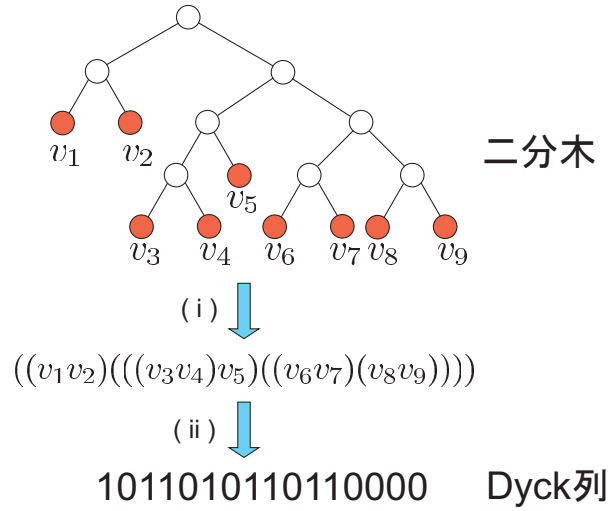


図 4.4 サイズ 9 の二分木から長さ 16 の Dyck 列を得る方法。(i) 二分木を葉  $\{v_1, \dots, v_9\}$  および  $()$  で表す。(ii)  $v_1$  および  $()$  を除去し、 $v_2, \dots, v_9$  を  $'1'$  に、 $'()$  を  $'0'$  に置き換える。

- (2) 列の中の  $'1'$  の総数が  $S_{1,n}(T)$  である。また、 $r \geq 2$  に対して  $S_{r,n}(T) = 0$  としておく。
- (3) 列の中に  $'m \ n \ 0'$  ( $m, n > 0$ ) という形の部分を 1 つの数  $'\max\{m, n\} + \delta_{m,n}'$  で置き換える (次数の計算と同じ計算)。  $'(r-1) \ (r-1) \ 0' \rightarrow 'r'$  という置き換えを行なうたびに  $S_{r,n}(T)$  を 1 ずつ増やす。
- (4) 列の長さが 1 になるまで繰り返す。

一連の計算手順を図 4.5 に示す。計算過程で二分木の構造が再現されている。

二分木の個数と Dyck 経路の個数がともに Catalan 数で表されることは古くからよく知られていて、二分木と Dyck 経路の間の 1 対 1 対応やそれに関連した種々の問題も議論されている [47, 56, 73–76]。ただし、逆ポーランド記法に基づく対応法を用いると二分木の次数の計算において図 4.5 に示した “より深い対応” が現れることへの言及はされていない。

## 4.4 ランダムな Dyck 経路の生成

Dyck 経路を等確率で生成する手法は [77–80] 等により調べられているが、本節ではこれらとは若干異なる方法を提案する。また、手法の視覚的な解釈も与える。

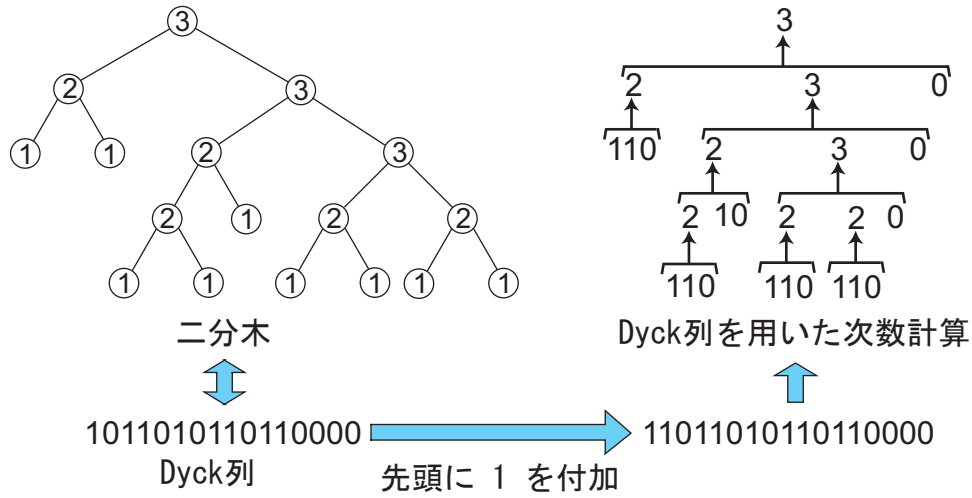


図 4.5 Dyck 列を用いて二分木の次数を計算する方法。計算過程で二分木の構造が再現されている。

Dyck 経路が通りうる格子点の全体を  $\mathcal{D}$  で表す。すなわち  $\mathcal{D} = \{(x, y) \in \mathbb{Z}^2 | 0 \leq x \leq n-1, 0 \leq y \leq n-1, x \geq y\}$  である。各点  $(x, y) \in \mathcal{D}$  に対して遷移確率  $P_{\uparrow}(x, y)$  および  $P_{\rightarrow}(x, y)$  を割り当てる。このとき、Dyck 経路  $(s_0, s_1, \dots, s_{2(n-1)})$  が生じる確率を

$$P(s_0, \dots, s_{2(n-1)}) = \prod_{i=0}^{2(n-1)-1} p_i, \quad \text{ただし } p_i = \begin{cases} P_{\rightarrow}(s_i), & (s_i, s_{i+1}) \text{ が右への遷移} \\ P_{\uparrow}(s_i), & (s_i, s_{i+1}) \text{ が上への遷移} \end{cases} \quad (4.1)$$

で定義する。すなわち、Dyck 経路を状態空間  $\mathcal{D}$ 、遷移確率  $\{P_{\rightarrow}(s), P_{\uparrow}(s)\}_{s \in \mathcal{D}}$  の Markov 連鎖とみなし、その生成確率を計算している。二分木のランダムモデルに対応させるには、全ての Dyck 経路の生成確率 (4.1) が一定になるように各格子点の遷移確率  $\{P_{\rightarrow}(s), P_{\uparrow}(s)\}_{s \in \mathcal{D}}$  を決定しなければならない。

点  $(x, y) \in \mathcal{D}$  から  $(n-1, n-1)$  への  $\mathcal{D}$  上の経路の個数を  $\nu(x, y)$  で表す。ちなみに、 $(x, y)$  から  $(n-1, n-1)$  への経路の長さは  $2(n-1) - (x+y)$  であり、 $(0, 0)$  と  $(n-1, n-1)$  を結ぶ  $\mathcal{D}$  上の経路は Dyck 経路にほかならない。 $\nu(x, y)$  の計算も反射原理を利用すれば簡単である。まず、 $(x, y)$  から  $(n-1, n-1)$  への全ての経路 ( $\mathcal{D}$  からはみ出してもよい) の個数は  $\binom{2(n-1)-(x+y)}{n-x-1}$  である ( $2(n-1) - (x+y)$  は格子路の長さ、 $n-x-1$  は右に進む回数)。次に、 $(x, y)$  から  $(n-1, n-1)$  への経路のうちで、 $\mathcal{D}$  からはみ出してしまうものの総数は反射原理によって  $(x, y)$  から  $(n-2, n)$  への経路数と等しく、 $\binom{2(n-1)-(x+y)}{n-x-2}$  で与

えられる。これより、

$$\begin{aligned}\nu(x, y) &= \binom{2(n-1) - (x+y)}{n-x-1} - \binom{2(n-1) - (x+y)}{n-x-2} \\ &= \frac{\{2(n-1) - (x+y)\}!}{(n-1-x)!(n-y)!} (x-y+1)\end{aligned}\quad (4.2)$$

と計算できる。この  $\nu(x, y)$  は Kostka 数と呼ばれる数え上げ関数の一種であり、Young 図形の数え上げなど組合せ論の他の問題にも現れる [46, 81]。

$\nu(x, y)$  は次の各性質を満たしている（いずれも (4.2) を用いて素直に計算すれば示せる）。

- (1) 全ての  $(x, y) \in \mathcal{D}$  について  $\nu(x, y) \geq 0$
- (2)  $\nu(n-1, y) = 1$  ただし  $y = 0, 1, \dots, n-1$
- (3)  $m = 0, 1, \dots, n-1$  に対して  $\nu(m, m) = \frac{\{2(n-m-1)\}!}{(n-m-1)!(n-m)!}$ 。つまり、 $\nu(m, m)$  は  $n-m-1$  番目の Catalan 数である。特に  $\nu(0, 0)$  は  $n-1$  番目の Catalan 数  $c_{n-1}$  である。
- (4) 全ての  $(x, y) \in \mathcal{D}$  に対して  $\nu(x, y) = \nu(x+1, y) + \nu(x, y+1)$  が成り立つ。ただし、 $(x+1, y)$  や  $(x, y+1)$  が  $\mathcal{D}$  からみ出している場合はその点での  $\nu$  の値を 0 とする。この関係式は  $(x, y)$  から  $(n-1, n-1)$  への全ての経路は最初の遷移が右か上のいずれかであることに対応する。

各点  $(x, y) \in \mathcal{D}$  に対して遷移確率  $P_{\rightarrow}(x, y)$  および  $P_{\uparrow}(x, y)$  を

$$\begin{aligned}P_{\rightarrow}(x, y) &= \frac{\nu(x+1, y)}{\nu(x, y)} = \frac{(n-1-x)(x-y+2)}{(1+x-y)\{2(n-1) - (x+y)\}}, \\ P_{\uparrow}(x, y) &= \frac{\nu(x, y+1)}{\nu(x, y)} = \frac{(n-y)(x-y)}{(1+x-y)\{2(n-1) - (x+y)\}}\end{aligned}\quad (4.3)$$

によって定義する。点  $(x+1, y)$  や  $(x, y+1)$  が  $\mathcal{D}$  上にない場合に  $\nu$  を 0 とするのは上と同様である。上の性質 (4) より  $P_{\rightarrow}(x, y) + P_{\uparrow}(x, y) = 1$  がただちに分かる。これは確率の規格化条件である。対角線上の格子点  $(m, m)$  においては  $P_{\uparrow}(m, m) = 0$  である。これは、対角線上の格子点は上への遷移が許されない（ $\mathcal{D}$  から飛び出してしまう）ことと一致している。また、右端の格子点  $(n-1, m)$  においては  $P_{\rightarrow}(n-1, m) = 0$  である。右端では上への遷移のみが可能だからである。

$\nu(x, y)$  は次の規則によって“手計算”することができる。

- (1)  $(x, y) \notin \mathcal{D}$  に対しては  $\nu(x, y) = 0$  とする。
- (2)  $\mathcal{D}$  の右端の格子点  $(n-1, y)$  ( $y = 0, 1, \dots, n-1$ ) では  $\nu(n-1, y) = 1$ 。

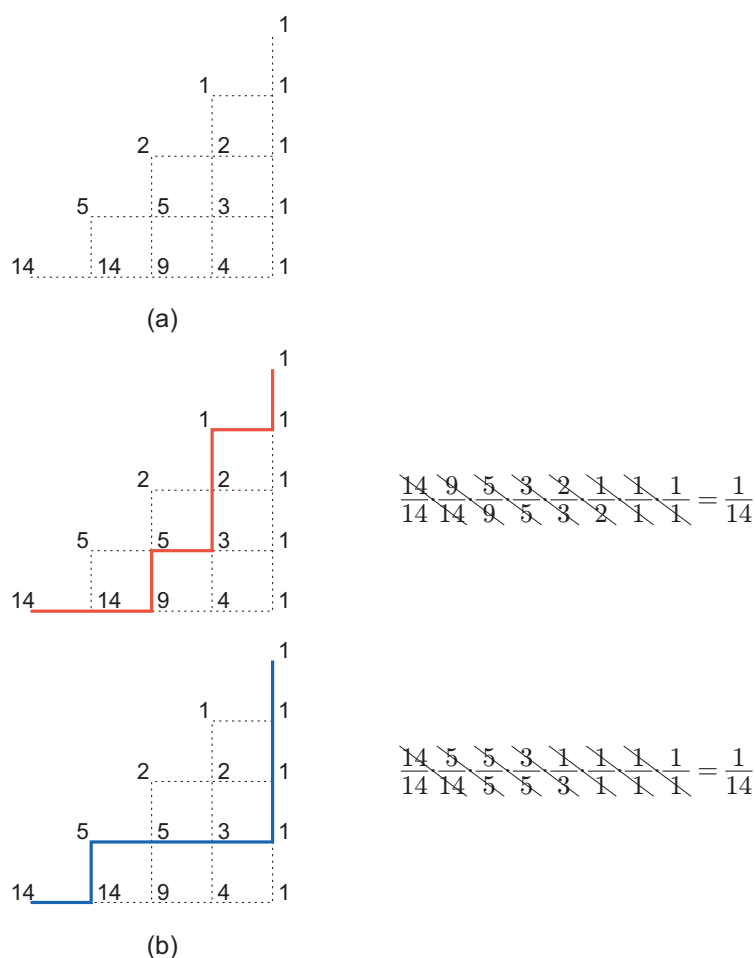


図 4.6 (a)  $\nu(x, y)$  の計算例 ( $n = 5$ )。 (b) 2 つの Dyck 経路とその生成確率の計算。分子と分母の打ち消し合いが続き、生成確率はともに  $\frac{1}{14}$  となる。

- (3)  $\mathcal{D}$  の他の格子点  $(x, y)$  に対しては関係式  $\nu(x, y) = \nu(x + 1, y) + \nu(x, y + 1)$  を利用して  $\nu(x, y)$  を求める。つまり、 $\nu(x, y)$  は  $(x, y)$  の右隣および上隣の点での  $\nu$  の値を合計して求められる。

Dyck 経路の生成確率は経路上の格子点に書かれた  $\nu(x, y)$  の値を分母・分子に書いていくことで計算できる (図 4.6(b) を参照)。経路の途中の点での  $\nu(x, y)$  は次々と約分されていくため、Dyck 経路の生成確率は

$$\frac{\nu(n-1, n-1)}{\nu(0, 0)} = \frac{1}{c_{n-1}}$$

で与えられ、全ての Dyck 経路が等確率で生成されることが分かる。



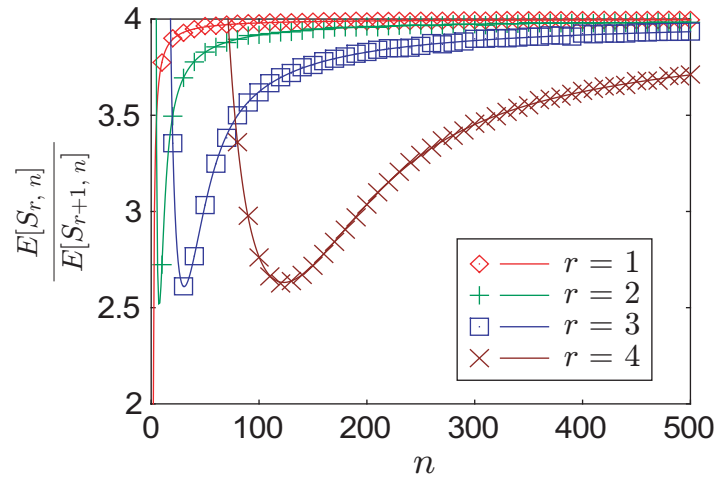


図 4.7 数値計算アルゴリズムの適用例。各点が本章のアルゴリズムを用いて計算したもの。それぞれの  $n$  ごとに  $10^5$  個のサンプルを平均した。曲線は第2章の漸化式 (2.8) を直接計算した結果である。比較しやすいようにアルゴリズムによる結果はサイズ  $n$  を 10 刻みに間引いて表示した。

## 4.5 ランダムモデルの数値計算アルゴリズム

前節までの結果を踏まえてランダムモデルの確率変数  $S_{r,n}$  に対する数値計算法を述べる。まず、Dyck 列を等確率で生成する方法を整理しておく。

- (1)  $\mathcal{D}$  上の位置を初期化する:  $(x, y) = (0, 0)$ . Dyck 列は空集合  $\emptyset$  に初期化する。
- (2) 式 (4.3) を用いて、位置  $(x, y)$  での遷移確率  $P_{\rightarrow}$  および  $P_{\uparrow}$  を計算し、この確率に応じて遷移する方向を選択する。
- (3) 右に遷移する場合は  $x$  に 1 を加え、Dyck 列の末尾に ‘1’ を追加する。上に遷移する場合は  $y$  に 1 を加え、Dyck 列の末尾に ‘0’ を追加する。
- (4)  $(x, y) = (n-1, n-1)$  になるまで (2), (3) を繰り返す。

次に、一般的な確率変数  $f(S_{r,n})$  の期待値  $E[f(S_{r,n})]$  を数値的に評価する手順を示す。

- (1) 上の方法により等確率の Dyck 列を生成する。
- (2) 図 4.5 にしたがって Dyck 列に対応する二分木の頂点の次数を計算し、 $S_{r,n}$  の値 (非負整数) を求める。
- (3)  $S_{r,n}$  の値から  $f(S_{r,n})$  を計算する。

(4) (1)-(3) を多数回繰り返す、得られた  $f(S_{r,n})$  の平均値を  $E[f(S_{r,n})]$  の計算結果とする。

$r = 1, 2, 3, 4$  の分岐比  $\frac{E[S_{r,n}]}{E[S_{r+1,n}]}$  の数値計算結果を漸化式 (2.8) を直接計算した結果とともに図 4.7 に示す。本節のアルゴリズムを用いた結果（プロット）は漸化式の計算の結果（曲線）とよく一致している。

## 第5章 二分木に対する中心極限定理

第3章までで一般的な確率変数  $f(S_{r,n})$  の期待値の漸近展開 (3.11) を得た。しかし、確率変数の性質は期待値だけでは記述しきれない。本章では逆分岐比という確率変数を新たに導入し、この確率変数に関する中心極限定理を数値的に検証する。前章で述べた数値計算法の適用例でもある。

### 5.1 はじめに – 中心極限定理と誤差関数

$(X_1, X_2, \dots)$  を同一の分布にしたがう独立な確率変数の列とし、平均  $\mu = E[X_1] < \infty$ , 分散  $\sigma^2 = \text{var}(X_1) < \infty$  とする。このとき、先頭の  $n$  個の和  $S_n = X_1 + \dots + X_n$  について

$$\sqrt{n} \left( \frac{S_n}{n} - \mu \right) \Rightarrow N(0, \sigma^2) \quad \text{as } n \rightarrow \infty \quad (5.1)$$

が成り立つというのが中心極限定理である。ただし、 $N(\mu, \sigma^2)$  は平均  $\mu$ , 分散  $\sigma^2$  の正規分布を表す。また、‘ $\Rightarrow$ ’ は分布収束を表し、 $S_n$  の分布関数の各点収束

$$P \left( \sqrt{n} \left( \frac{S_n}{n} - \mu \right) \leq x \right) \rightarrow \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{\xi^2}{2\sigma^2}} d\xi \quad \text{for } \forall x \in \mathbb{R}$$

を意味する。確率変数の収束については、互いに等価でない概念が数種類知られているが、分布収束はその中でも最も弱い [82]。中心極限定理は通常、左辺の特性関数が正規分布  $N(0, \sigma^2)$  の特性関数に各点収束することから証明される [83]。また、 $X_1, X_2, \dots$  が同一の分布とは限らない場合には例えば、Lindeberg の中心極限定理 [82] などの変形が知られている。

標準正規分布  $N(0, 1)$  の累積密度関数 (分布関数)  $F(x)$  は

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi + \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi = \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{x}{\sqrt{2}} \right)$$

で与えられる (図 5.1)。ただし、 $\text{erf}(x)$  は (Gauss の) 誤差関数で、

$$\text{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi \quad (5.2)$$

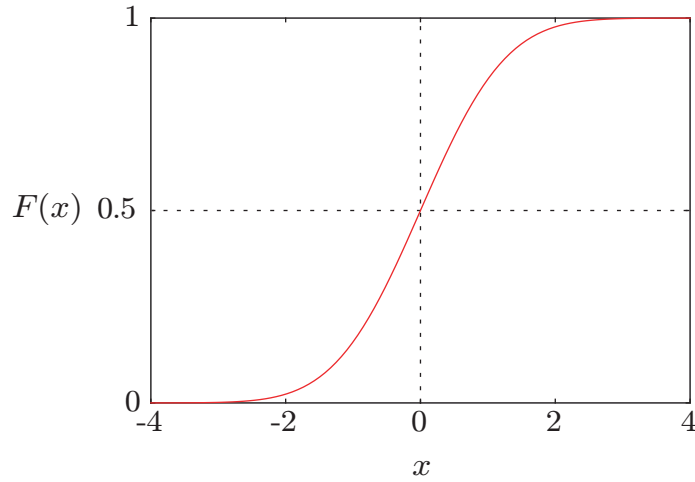


図 5.1 標準正規分布  $N(0, 1)$  の累積密度関数  $F(x)$  のグラフ。

により定義される。対称性をよくするために (5.2) の右辺にある積分の下端が 0 であることに注意 (累積密度関数の積分の下端は  $-\infty$  である)。 $\operatorname{erf}(x)$  は  $\mathbb{R}$  上狭義単調増加な奇関数であり、無限遠では  $\lim_{x \rightarrow \pm\infty} \operatorname{erf}(x) = \pm 1$  に収束する (無限遠での値の計算は Gauss 積分にほかならない)。積分の変数変換により、一般の正規分布  $N(\mu, \sigma^2)$  の累積密度関数  $F_{\mu, \sigma}(x)$  は

$$F_{\mu, \sigma}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) \quad (5.3)$$

と計算できる。 $F_{\mu, \sigma}(x)$  は  $F(x)$  を  $x$  方向に並進・スケール変換することで得られる。

## 5.2 $S_{2,n}$ に対する中心極限定理とその拡張

Wang and Waymire [84] によると、 $\Omega_n$  上の確率変数  $S_{2,n}$  に対して次の形の中心極限定理

$$\sqrt{n} \left( \frac{S_{2,n}}{n} - \frac{1}{4} \right) \Rightarrow N \left( 0, \frac{1}{16} \right), \quad \text{as } n \rightarrow \infty \quad (5.4)$$

が成り立つことが証明されている。式の形自体は (5.1) とよく似ているが、ランダムモデルの確率変数  $S_{2,n}$  は独立な確率変数の和で表すことができないという違いがある。そのため、通常を中心極限定理の証明をそのまま流用することはできない。実際、[84] の証明は母関数の計算から得た大偏差原理の系として中心極限定理を示すという規模の大きなものである。

本章では定理 (5.4) の単純な拡張を 2 種類提案する。まず、最も素直な一般化が (5.4) の左辺の  $S_{2,n}$  を  $S_{r,n}$  に置き換えることであろう。式 (2.14) や (3.14) より

$$\lim_{n \rightarrow \infty} \frac{E[S_{r,n}]}{n} = \frac{1}{4^{r-1}}$$

となるため、

$$\sqrt{n} \left( \frac{S_{r+1,n}}{n} - \frac{1}{4^r} \right) \Rightarrow N(0, \sigma_r^2), \quad \text{as } n \rightarrow \infty \quad (5.5)$$

と表せる。ただし、 $\sigma_r^2$  は未知の分散であり、一般に次数  $r$  に依存する。また、推測で立てた式であり正当性が保証されていないという意味で数式を “...” で囲んでいる。

Horton-Strahler 解析では分岐比が重要な量であり、期待値の比  $E[S_{r,n}]/E[S_{r+1,n}]$  として定義していた。単純に考えれば  $S_{r,n}/S_{r+1,n}$  という確率変数を分岐比と定義してもよさそうである。任意の  $n \in \mathbb{N}$ 、全ての二分木  $T \in \Omega_n$  に対して  $S_{r,n}(T)$  は次数  $r$  について単調減少であり（高次の枝ほど少ないということ）、 $r$  が充分大きければ  $S_{r,n}(T) = 0$  となる（具体的には、 $r \geq 2 + \log_2 n$  であればよい）。よって、 $S_{\tilde{r},n}(T) \neq 0$  かつ  $S_{\tilde{r}+1,n} = 0$  となる  $\tilde{r}$  が必ず存在する。 $S_{\tilde{r},n}(T)/S_{\tilde{r}+1,n}(T)$  は定義できない（正の数を 0 で割っている）。さらに、この  $\tilde{r}$  は  $T$  に依存するため、多くの  $r$ （具体的には  $2 \leq r \leq \lfloor \log_2 n \rfloor + 1$ ）について確率変数  $S_{r,n}/S_{r+1,n}$  が  $\Omega_n$  全体で定義できない。したがって、確率変数  $S_{r,n}/S_{r+1,n}$  を扱うのは面倒である。

その代わり、本章では  $Q_{r,n} := S_{r+1,n}/S_{r,n}$  という確率変数を取り扱う。 $T \in \Omega_n$  に対して確率変数の値を

$$Q_{r,n}(T) = \begin{cases} \frac{S_{r+1,n}(T)}{S_{r,n}(T)} & r \leq \tilde{r} \\ 0 & r \geq \tilde{r} \end{cases}$$

と定義する。 $r = \tilde{r}$  でも矛盾なく定義されていることに注意。この確率変数  $S_{r+1,n}/S_{r,n}$  を“分岐比の逆数に相当する確率変数”という意味で逆分岐比と呼ぶことにする。(3.16) より、 $n \rightarrow \infty$  での  $Q_{r,n}$  の期待値は

$$\lim_{n \rightarrow \infty} E[Q_{r,n}] = \frac{1}{4}$$

である。

$S_{1,n} \equiv n$  であるため、(5.4) は  $Q_{1,n}$  に関する極限定理

$$\sqrt{n} \left( Q_{1,n} - \frac{1}{4} \right) \Rightarrow N \left( 0, \frac{1}{16} \right)$$

の形に書ける。これを素直に一般の  $r$  へ拡張すると

$$\sqrt{n} \left( Q_{r,n} - \frac{1}{4} \right) \Rightarrow N(0, \tilde{\sigma}_r^2) \quad (5.6)$$

となる。

次節では2種類の中心極限定理(5.5), (5.6)を仮定し、次数  $r$  に依存した分散  $\sigma_r^2$  および  $\tilde{\sigma}_r^2$  を数値的に評価する。定理(5.4)を一般の次数に拡張する方法は何通りも考えられる。しかし、最も単純なやり方が本節で提案した2つであろう。第1の式(5.5)は  $S_{r,n}$  のふるまい、第2の式(5.6)は逆分岐比のふるまいに関する主張であり、どちらも重要な意味をもっている。

### 5.3 数値計算結果

数値計算においてサイズ  $n$  を 10000, 反復回数を  $10^5$  とした。

- (1) 前章のアルゴリズムで二分木  $T$  を等確率で生成し、各々の  $r$  について  $S_{r,n}(T)$  を計算する。
- (2) それぞれの  $r$  について  $\sqrt{n} \left( \frac{S_{r+1,n}(T)}{n} - \frac{1}{4^r} \right)$  および  $\sqrt{n} \left( \frac{S_{r+1,n}(T)}{S_{r,n}(T)} - \frac{1}{4} \right)$  を計算
- (3) (1),(2) を  $10^5$  回繰り返す。
- (4) 得られたデータから累積密度関数を計算。
- (5) 関数(5.3)のフィッティングから分散  $\sigma_r^2, \tilde{\sigma}_r^2$  を決定。

図5.2が数値計算の結果である。どのデータも誤差関数のグラフ(図5.1)と似ている。確率変数  $\sqrt{n} \left( \frac{S_{r+1,n}}{n} - \frac{1}{4^r} \right)$  については、次数  $r$  が大きいほど分散は小さくなっている。一方、 $\sqrt{n} \left( \frac{S_{r+1,n}}{S_{r,n}} - \frac{1}{4} \right)$  は次数  $r$  が大きいほど分散は大きくなっている。また、(b)において  $r=5$  および  $r=6$  のグラフが階段状になっているのは、 $r$  が大きいほど分母  $S_{r,n}(T)$  は小さくなり、 $S_{r+1,n}(T)/S_{r,n}(T)$  のとりうる値の間隔が大きくなってしまいうためである。次節でこれらの結果の検討を行なう。

### 5.4 結果の検討

本節では数値計算の結果と理論的な計算の比較を行なう。まず、3.4節で得た(3.15)を用いると、

$$\sigma_r^2 = \lim_{n \rightarrow \infty} \text{var} \left( \sqrt{n} \cdot \frac{S_{r+1,n}}{n} \right) = \lim_{n \rightarrow \infty} \frac{\text{var}(S_{r+1,n})}{n} = \frac{1}{3} \frac{4^r - 1}{16^r} \quad (5.7)$$

が得られる。

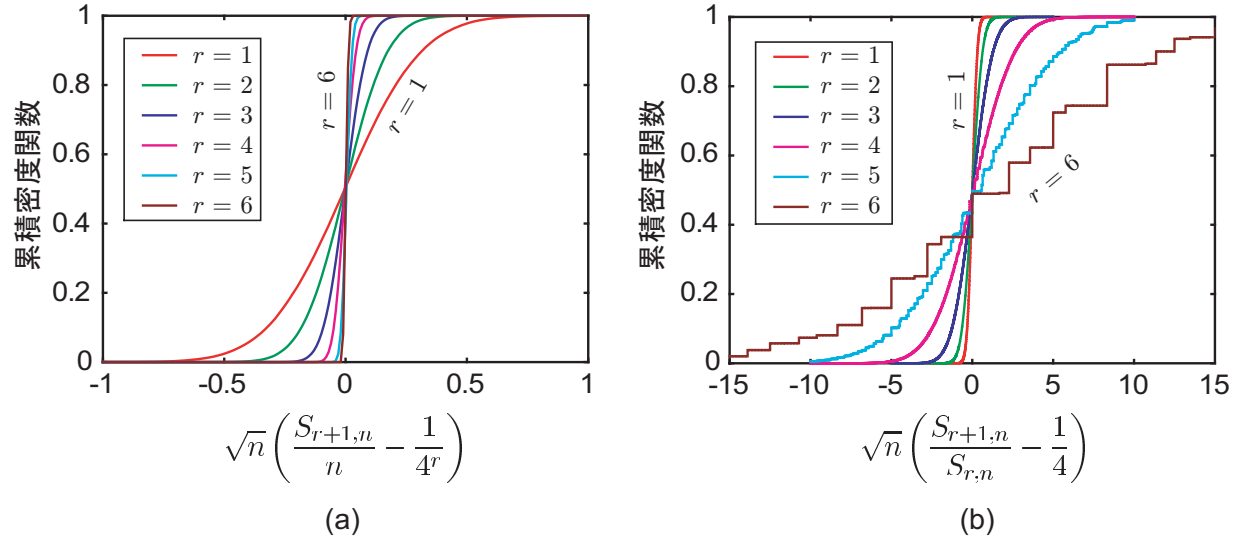


図 5.2 数値データから構成した累積密度関数 ( $n = 10000$ , サンプル数  $10^5$ )。 (a) 確率変数  $\sqrt{n} \left( \frac{S_{r+1,n}}{n} - \frac{1}{4^r} \right)$  の累積密度関数。次数  $r$  が大きいほど分散  $\sigma_r^2$  は小さい。 (b) 確率変数  $\sqrt{n} \left( \frac{S_{r+1,n}}{S_{r,n}} - \frac{1}{4} \right)$  の累積密度関数。次数  $r$  が大きいほど分散  $\tilde{\sigma}_r^2$  が大きい。

次に、逆分岐比に関する分散  $\tilde{\sigma}_r^2$  を調べる。 $S_{2,n}$  の分散の漸近展開 (3.15) で  $r = 2$  の場合を用いると、1 次の逆分岐比  $Q_{1,n} = S_{2,n}/S_{1,n}$  の分散は

$$\text{var} \left( \frac{S_{2,n}}{S_{1,n}} \right) = \text{var} \left( \frac{S_{2,n}}{n} \right) = \frac{\text{var}(S_{2,n})}{n^2} = \frac{1}{16n} - \frac{1}{32n^2} + O(n^{-3})$$

と展開される。よって、 $a_1 = \frac{1}{16}$ ,  $b_1 = -\frac{1}{32}$ , および  $k = -1$  を公式 (3.12) に代入すると

$$\text{var} \left( \frac{S_{r+1,n}}{S_{r,n}} \right) = \left( \frac{n}{4^{r-1}} \right)^{-1} \left\{ \frac{1}{16} + \frac{1}{n} \left( \frac{4^{r-1}}{48} + \frac{1}{96} \right) \right\} + O(n^{-3})$$

$n \rightarrow \infty$  の極限をとると

$$\tilde{\sigma}_r^2 = \lim_{n \rightarrow \infty} \text{var} \left( \sqrt{n} \cdot \frac{S_{r+1,n}}{S_{r,n}} \right) = \lim_{n \rightarrow \infty} n \cdot \text{var} \left( \frac{S_{r+1,n}}{S_{r,n}} \right) = 4^{r-1} \cdot \frac{1}{16} = 4^{r-3} \quad (5.8)$$

が得られる。

表 5.1 に標準偏差  $\sigma_r$ ,  $\tilde{\sigma}_r$  のフィッティングから得られた値と理論式 (5.7), (5.8) から計算された値を示す。両者はよく一致している。また、 $\sigma_r$  と  $\tilde{\sigma}_r$  を  $r$  の関数としてプロットしたのが図 5.3 である。図中の理論曲線は

$$\sigma_r = \frac{1}{4^r} \sqrt{\frac{4^r - 1}{3}}, \quad \tilde{\sigma}_r = 2^{r-3}$$

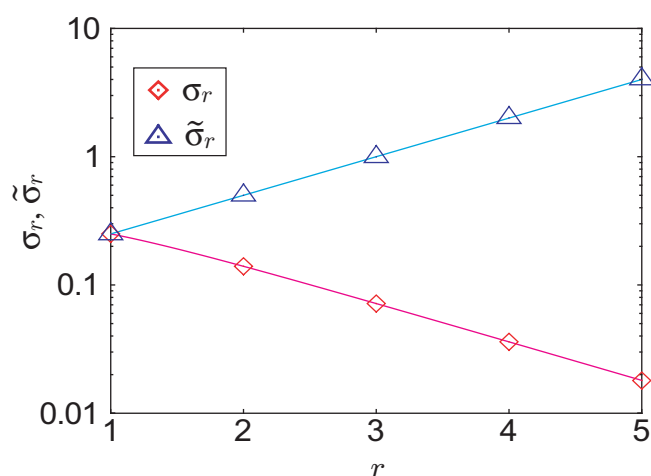


図 5.3  $r$  の関数としての  $\sigma_r$ ,  $\tilde{\sigma}_r$  の振る舞い。プロットは数値データのフィッティング結果、曲線は理論式を表す。

である。

表 5.1 2 種類の標準偏差  $\sigma_r$ ,  $\tilde{\sigma}_r$  の数値計算結果と理論値。 $\sigma_r$ ,  $\tilde{\sigma}_r$  の理論値はそれぞれ (5.7), (5.8) の平方根をとったものである。

$r$	$\sigma_r$		$\tilde{\sigma}_r$	
	数値計算	理論値 (5.7)	数値計算	理論値 (5.8)
1	$0.2502 \pm 0.0004$	0.25	$0.2492 \pm 0.0001$	0.25
2	$0.1397 \pm 0.0003$	0.1398	$0.5000 \pm 0.0001$	0.5
3	$0.0716 \pm 0.0001$	0.0716	$0.9967 \pm 0.0004$	1
4	$0.0361 \pm 0.0004$	0.0360	$2.0085 \pm 0.0003$	2
5	$0.0181 \pm 0.0009$	0.0180	$4.025 \pm 0.002$	4

以上より、 $r = 1, 2, \dots$  に対して

$$\begin{aligned} \sqrt{n} \left( \frac{S_{r+1,n}}{n} - \frac{1}{4^r} \right) &\Rightarrow N \left( 0, \frac{1}{3} \frac{4^r - 1}{16^r} \right), \\ \sqrt{n} \left( \frac{S_{r+1,n}}{S_{r,n}} - \frac{1}{4} \right) &\Rightarrow N \left( 0, 4^{r-3} \right) \end{aligned} \quad (5.9)$$

という 2 種類の中心極限定理が予想される。 $r = 1$  の場合はどちらも厳密な結果 (5.4) に一致している。

図 5.2 および表 5.1 をみる限り、中心極限定理 (5.9) は非常にもっともらしい。しかし、これらの結果は数値的に得られたものであり、はじめに仮定した“正規分布への収束”など、



厳密に言えば数学的に明らかでないことばかりである。ただし、本章の結果は拡張した表現の中心極限定理 (5.9) を予想し、解析的な議論へ向けた動機付け・指針を与えたという意味をもっている。

## 第6章 総括

本論文では分岐系の確率モデルの1つであるランダムモデル上での確率変数  $S_{r,n}$  の漸近的な性質を議論した。

第2章では  $S_{r,n} = n_r$  であるような二分木  $T$  の総数を求めるために完全二分木を順次ぶら下げていく方法を述べた。基本となるのが“ $r$  次の枝が生じるためには  $r$ -完全二分木が埋め込まれていなければならない”という性質である。特に、 $S_{2,n}$  については様々な量が厳密に計算可能であり、モーメント、分散等の表式を特殊関数の公式を駆使して導出した。これらの中には先行研究で得られていた結果の再導出や一般化にあたるものも含まれていた。一方  $r \geq 3$  のとき、 $S_{r,n}$  の期待値やモーメント等の厳密な表示を得るのは困難である（おそらく既知の特殊関数では表現できないであろう）。代わりに  $S_{r,n}$  のモーメントが満たす漸化式を導出し、これを用いて  $S_{r,n}$  のモーメントを数値的に評価することができた。ただし、 $r$  が大きいと計算コストが大きくなりすぎてしまう。 $n$  が十分に大きい場合には漸化式を解くことができ、 $S_{r,n}$  のモーメントの漸近形が得られた。これによりランダムモデルが漸近的に分岐比 4 の位相的自己相似性をもつことが示された。

第3章では  $f(S_{r,n})$  という一般的な形の確率変数の性質を議論した。ただし、関数  $f$  は複素関数とみた場合に無限遠点で Laurent 展開が可能であるという比較的ゆるい仮定が必要だった。第2章で  $S_{r,n}$  の期待値が解析的に表せないことが示唆されたため、当然  $f(S_{r,n})$  の期待値も有限の  $n$  では厳密に計算することはできない。 $E[f(S_{r,n})]$  の漸近形を求めるために、第2章で得たものと類似の漸化式を導出した。その導出のために二分木の粗視化に相当する変換を導入した。そのおかげで第2章よりも洗練された形で漸化式を得ることができた。 $n$  が大きいときの漸化式の解として  $f(S_{r,n})$  の漸近展開の公式を得た。漸近展開の公式を用いると  $n \rightarrow \infty$  では  $E[f(S_{r,n})]$  と  $E[f(S_{r+1,n})]$  の比が  $r$  によらないことが簡単に示される。これはランダムモデルがより強い位相的自己相似性をもつことを表している。さらに、いくつかの重要な計算例も示した。

第4章ではランダムモデルの数値計算法を提案した。その内容は、(1) 二分木を Dyck 列というビット列に変換する、(2) Dyck 列を用いて対応する二分木の各頂点の次数を計算する、(3) Dyck 列を等確率で生成する、の3点から構成されている。二分木と Dyck 列の対応付けは古典的な問題であり何通りかのやり方が知られている。本論文では二項演算の逆ポーランド記法に基づく方法を採用し、Dyck 列と二分木が次数の計算において深い対応

関係がみられることを明らかにした。また、格子上の Markov 過程の枠組みで Dyck 経路を等確率で生成する方法を述べた。 $E[S_{r,n}]$  の数値計算と第2章の漸化式による結果を比較し、アルゴリズムの有効性を確認した。

第5章では  $S_{r,n}$  に関する中心極限定理を定式化した。 $S_{2,n}$  に関する中心極限定理は解析的に証明がなされているのだが、本章でこの結果を  $S_{r,n}$  に関する主張に一般化することを試みた。単純かつ意味のある拡張が2種類考えることができる。数値計算および分散の極限値の理論計算から2種類の中心極限定理が示唆された。これらは厳密な証明が与えられてはいないが、今後の解析的な証明へつながる結果であると期待している。

本論文でランダムモデルでの  $S_{r,n}$  の挙動に対する理解が深まったが、完全に理解できたわけではない。例えば中心極限定理やその精密化にあたる大偏差原理などの極限定理は  $S_{r,n}$  の漸近的な性質を知るうえで基本的かつ重要であるが厳密な結果が知られているのは  $r = 2$  の場合のみである。また、アンサンブル理論等を用いた特徴付けやパーコレーションや繰り込みなど統計力学と関連の深い性質の解析も物理との接点において不可欠である。これらは問題としての定式化自体なされていないのが現状だが、ランダムモデルの数学的な単純さを生かせば厳密で自明でない結果も十分に望める。さらに、ランダムモデルで得られた結果が現実の分岐パターンでも成り立つのかを検証することも非常に重要である。この作業によって分岐パターンのもつ新たな性質や多くの分岐パターンに共通して現れる普遍性が明らかになるかもしれない。

## 謝辞

本学位論文を完成させるにあたり、指導教官である山崎義弘先生には研究全般にわたってご指導いただき、たいへんお世話になりました。身のまわりにある些細な事象も突き詰めれば自然科学の研究対象になるという精神を学ばせていただきました。

山崎研究室の助手の小村真也さんには、度々にわたり数値計算に関する適切な助言をいただきました。私が最低限のプログラミング技術を何とか身につけることができたのは小村さんが開発された HML のおかげです。

相澤洋二先生、大谷光春先生、栗原進先生ならびに中央大学理工学部の松下貢先生には博士論文審査の副査をご担当いただきました。公聴会では今後の研究の方向性なども含めて広い視野からご指摘・ご助言があり、非常にありがたく感じております。

松下先生からは毎年末の研究発表会でも有益なコメントをいただきました。また投稿前の論文に目を通していただき、ありがとうございました。

日本学術振興会からは研究費をご支援いただきまして、たいへん感謝しております。

なお、先進理工学部物理学科 1 年次配当の科目「物理学研究ゼミナール」(2007 年度) が本研究の発端となっております。山崎研究室の担当になった当時の 1 年生にもささやかな謝意を送りたいと思います。

最後に、博士課程まで物理を学ばせてくれたうえ、経済的・精神的・栄養的にサポートしてくれた両親および家族に書き尽くせない感謝の意を捧げ、謝辞といたします。

## 研究業績

### 論文

- (1) K. Yamamoto and Y. Yamazaki, Central limit theorem of the bifurcation ratio of the Horton-Strahler analysis: application to DLA clusters, *J. Phys. Soc. Jpn.* **80**, to be published.
- (2) K. Yamamoto and Y. Yamazaki, Topological self-similarity on the random binary-tree model. *J. Stat. Phys.* **139**, 62-71 (2010).
- (3) K. Yamamoto and Y. Yamazaki, Central limit theorem for bifurcation ratio of random binary tree. *J. Phys. A: Math. Theor.* **42**, 415002 (2009).
- (4) K. Yamamoto and Y. Yamazaki, Formulation and asymptotic properties of the bifurcation ratio in Horton's law for the equiprobable binary tree model. *Phys. Rev. E* **78**, 021114 (2008).

### 国内学会

- (1) 山本健, 小村真也, 山崎義弘, “粉粒体の乾燥過程で生じる迷路状パターンの分岐解析”, 日本物理学会 2010 年秋季大会, 大阪府立大学 (2010).
- (2) 山本健, 山崎義弘, “二分木の等確率モデルにおける漸化式と漸近的性質”, 日本物理学会 2008 年秋季大会, 岩手大学 (2008).

## 参考文献

- [1] V. Fleury, J.-F. Gouyet, and M. Léonetti, *Branching in Nature* (Springer, Berlin, 2001).
- [2] J.A. Davies, *Branching Morphogenesis* (Springer, Berlin, 2006).
- [3] P. Ball, *The Self-Made Tapestry* (Oxford University Press, Oxford, 1999).
- [4] P. Ball, *Branches* (Oxford University Press, Oxford, 2009).
- [5] 芸術工学基礎論への招待: <http://www/kobe-du.ac.jp/gsdr/gsdr/kiso04>
- [6] 丸安隆和, 八十島義乃助 監修, 新編土木工学ポケットブック (オーム社, 東京, 1982).
- [7] 国土交通省国土技術政策総合研究所, 河川用語集 “河川・本川・支川” の項:  
<http://www.nilim.go.jp/lab/rcg/newhp/yougo/words/009/009.html>
- [8] 土木学会 編, 土木用語大辞典 (技報堂, 東京, 1999).
- [9] N. Deo, *Graph Theory with Applications to Engineering and Computer Science* (Prentice-Hall, London, 1974).
- [10] D. Knuth, *The Art of Computer Programming, vol. 1* (Addison-Wesley, Reading, 1968).
- [11] R.E. Horton, Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol. Soc. Amer. Bull.* **56**, 275-370 (1945).
- [12] A.N. Strahler, Hypsometric (area-altitude) analysis of erosional topology. *Geol. Soc. Amer. Bull.* **63**, 1117-1142 (1952).
- [13] A.N. Strahler, Quantitative analysis of watershed geomorphology. *Trans. Amer. Geophys. Un.* **8**, 913-920 (1957).

- [14] E. Tokunaga, Consideration on the composition of drainage networks and their evolution. *Geogr. Rep. Tokyo Metrop. Univ.* **13**, 1-27 (1978).
- [15] R.L. Shreve, Infinite topologically random channel networks. *J. Geol.* **75**, 178-186 (1967).
- [16] E.R. Weibel, *The Morphometry of the Human Lung* (Springer, Berlin, 1963).
- [17] L.E. Milton and C.D. Ollier, A code for labeling streams, basins, and junctions in a drainage net. *J. Hydrol.* **3**, 66-68 (1965).
- [18] G. Ranalli and A.E. Scheidegger, Topological significance of stream labeling methods. *Bulletin of the International Association of Scientific Hydrology* **13**, 77-85 (1968).
- [19] B.B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman, San Francisco, 1982).
- [20] 國上真章, 名古屋大学フラクタル研究会内報, 1984; 高安秀樹, フラクタル (朝倉, 東京, 1986).
- [21] D.R. Morse, J.H. Lawton, M.M. Dodson, and M.H. Williamson, Fractal dimension of vegetation and the distribution of arthropod body lengths. *Nature* **314**, 731-733 (1985).
- [22] J. Tatsumi, A. Yamauchi, and Y. Kono, Fractal analysis of plant root systems. *Ann. Botany* **64**, 499-503 (1989).
- [23] P. La Barbera and R. Rosso, Fractal geometry of river networks. *Eos Trans. Am. Geophys. Un.* **68**, 1276 (1987).
- [24] D.J. Schuller, A.R. Rao, and G.D. Jeong, Fractal characteristics of dense stream networks. *J. Hydrol.* **243**, 1-16 (2001).
- [25] T.C. Halsey, Diffusion-limited aggregation: a model for pattern formation. *Phys. Today* **53**, 36-41 (2000).
- [26] A.H. Reis, Constructal view of scaling laws of river basins. *Geomorphology* **78**, 201-206 (2006).
- [27] 高木隆司, 形の数理 (朝倉, 東京, 1992).

- [28] H. Maître and M. Pinciroli, Fractal characterization of a hydrological basin using SAR satellite images. *IEEE Trans. Geoscience and Remote Sensing* **37**, 175-181 (1999).
- [29] J.D. Pelletier and D.L. Turcotte, Shapes of river networks and leaves: are they statistically similar? *Phil. Trans. R. Soc. Lond. B* **335**, 307-311 (2000).
- [30] B. Zhang and R.S. Govindaraju, Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *J. Hydrol.* **273**, 18-34 (2003).
- [31] L.B. Leopold, Trees and streams: the efficiency of branching patterns. *J. Theor. Biol.* **31**, 339-354 (1971).
- [32] S.T.A. Pickett and J.S. Kempf, Branching patterns in forest shrubs and understory trees in relation to habitat. *New Phytologist* **86**, 219-228 (1980).
- [33] N.D. Mitchell, M.R. Dardeau, and W.W. Schroeder, Colony morphology, age structure, and relative growth of two gorgonian corals, *Leptogorgia hebes* (Verrill) and *Leptogorgia virgulata* (Lamarck), from the northern Gulf of Mexico. *Coral Reefs* **12**, 65-70 (1993).
- [34] E.R. Abraham, The fractal branching of an arborescent sponge. *Marine Biol.* **138**, 503-510 (2001).
- [35] K. Horsfield, Morphometry of the small pulmonary arteries in man. *Circulation Research* **42**, 593-597 (1978).
- [36] E. VanBavel and J.A. Spaan, Branching patterns in the porcine coronary arterial tree. Estimation of flow heterogeneity. *Circulation Research* **71**, 1200-1212 (1992).
- [37] Z.V. Djordjevic, X. Feng Li, Won Soo Shin, S.L. Wunder, and G.R. Baran, Fractal and topological characterization of branching patterns on the fracture surface of cross-linked dimethacrylate resins. *J. Materials Sci.* **30**, 2968-2980 (1995).
- [38] 小野耕平, 田坂聡明, 上飯坂實, ホーтонаの法則による林道分岐過程の解析. *日本林學會誌* **73**, 89-97 (1991).
- [39] P. Ossadnik, Branch order and ramification analysis of large diffusion-limited-aggregation clusters. *Phys. Rev. A* **45**, 1058-1066 (1992).
- [40] A. Arenas, L. Danon, A. Díaz-Guilera, P.M. Gleise, and R. Guimerà, Community analysis in social networks. *Eur. Phys. J. B* **38**, 373-380 (2004).



- [41] R. Diestel, *Graph Theory* (Springer, New York, 2000).
- [42] A. Cayley, On the analytical forms called trees. *Philos. Mag.* **28**, 374-378 (1858).
- [43] J. Matoušek and J. Nešetřil, *Invitation to Discrete Mathematics* (Oxford University Press, Oxford, 2008).
- [44] 根上生也, 離散構造 (共立, 東京, 1993).
- [45] 山田裕史, 組合せ論プロムナード (日本評論社, 東京, 2009).
- [46] R.P. Stanley, *Enumerative Combinatorics, vol. 2* (Cambridge University Press, New York, 1999).
- [47] J.H. Conway and R.K. Guy, *The Book of Numbers* (Copernicus, New York, 1996).
- [48] R.L. Shreve, Statistical law of stream numbers. *J. Geol.* **74**, 17-37 (1966).
- [49] C. Werner, Two models for Horton's law of stream numbers. *Canadian Geographer* **16**, 50-68 (1972).
- [50] J.W. Moon, On Horton's law for random channel networks. *Ann. Discrete Math.* **8**, 117-121 (1980).
- [51] Z. Toroczkai, Topological classification of binary tree using the Horton-Strahler index. *Phys. Rev. E* **65**, 016130 (2001).
- [52] 徳永英二, 組合せ理論による水路数の法則の考察. *地理学評論* **47**, 696-707 (1974).
- [53] A. Meir, J.W. Moon, and J.R. Ponder, On the order of random channel networks. *SIAM J. Alg. Discr. Meth.* **1**, 25-33 (1980).
- [54] F. Ruskey, On the average shape of binary trees. *SIAM J. Alg. Discr. Meth.* **1**, 43-50 (1980).
- [55] V.K. Gupta and E. Waymire, On the formulation on an analytical approach to hydrologic response and similarity at the basin scale. *J. Hydrol* **65**, 95-123 (1983).
- [56] X.G. Viennot, Trees everywhere. *Proceedings of the 15th colloquium on CAAP'90*, 18-41 (1990).
- [57] L. Devroye and P. Kruszewski, A note on the Horton-Strahler number for random trees. *Inform. Process. Lett.* **52**, 155-159 (1994).

- [58] P. Flajolet, J.C. Raoult, and J. Vuillemin, The number of registers required for evaluating arithmetic expressions. *Theor. Comput. Sci.* **9**, 99-125 (1979).
- [59] R. Kemp, The average number of registers needed to evaluate a binary tree optimally. *Acta Informatica* **56**, 275-370 (1979).
- [60] I. Yekutieli and B.B. Mandelbrot, Horton-Strahler ordering of random binary trees. *J. Phys. A: Math. Gen.* **27**, 285-293 (1994).
- [61] H. Prodinger, On a problem of Yekutieli and Mandelbrot about the bifurcation ratio of binary trees. *Theor. Comput. Sci.* **181**, 181-194 (1997).
- [62] M.E. Nebel, On the Horton-Strahler number for combinatorica tries. *Theor. Inform. Appl.* **34**, 279-296 (2000).
- [63] J.H.M. Wedderburn, The functional equation  $g(x^2) = 2\alpha x + [g(x)]^2$ . *Ann. Math.* **24**, 121-140 (1922).
- [64] 森口繁一, 宇田川 久, 一松信, 岩波数学公式 III (岩波, 東京, 1987).
- [65] O. Kempthorne, L. Folks, *Probability, Statistics and Data Analysis* (Iowa State University Press, Ames, 1971).
- [66] 金丸昭治, 高棹琢馬, 水文学 (朝倉, 東京, 1975).
- [67] D. Stanton and D. White, *Constructive combinatorics* (Springer, New York, 1986).
- [68] 成嶋弘, 数え上げ組合せ論入門 (日本評論社, 東京, 1996).
- [69] K.J. Berkling, A computing machine based on tree structures. *IEEE Trans. Comput.* **20**, 404-418 (1971).
- [70] S. Okamura, I. Matushima, and Y. Yano, The effective learning support strategy for self learning with problem-based learning. *Proceedings of the 3rd International Conference on Creating, Connecting and Collaborating through Computing*, 150-157 (2005).
- [71] 富浦洋一, 渡辺日出雄, ここまできた自然言語処理 – 例文の収集とその利用 –. *情報処理* **41**, 762-796 (2000).
- [72] J.A. Ball, *Algorithms for RPN Calculators* (John Wiley & Sons, New York, 1978).

- [73] E. Deutsch, A bijection on Dyck paths and its consequences. *Discr. Math.* **179**, 253-256 (1998).
- [74] X.G. Viennot, A Strahler bijection between Dyck paths and planar trees. *Discr. Math.* **246**, 317-329 (2002).
- [75] D. Callan, Some bijections and identities for the Catalan and Fine numbers. *Séminaire Lotharingien de Combinatoire* **53**, Article B53e (2006).
- [76] N.S.S. Gu, N.Y. Li, and T. Mansour, 2-binary trees: bijections and related issues. *Discr. Math.* **308**, 1209-1221 (2008).
- [77] E. Mäkinen, Generating random binary trees – A survey. *Inform. Sci.* **115**, 123-136 (1999).
- [78] E. Barcucci, A.D. Lungo, and E. Pergola, Random generation of trees and other combinatorial objects. *Theor. Comput. Sci.* **218**, 219-232 (1999).
- [79] L. Alonso and R. Schott, *Random Generation of Trees: Random Generators in Computer Science* (Springer, Berlin, 1995).
- [80] B. Johnsen, Generating binary trees with uniform probability. *BIT Numerical Math.* **31**, 15 (1991).
- [81] B.E. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions* (Wadsworth & Books, Pacific Grove, 1991).
- [82] 伊藤清, 確率論 II (岩波, 東京, 1988).
- [83] R. Durrett, *Probability: Theory and Examples* (Duxbury Press, Belmont, 1996).
- [84] S.X. Wang and E.C. Waymire, A large deviation rate and central limit theorem for Horton ratios. *SIAM J. Discr. Math.* **4**, 575-588 (1991).