

Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України

Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

Богун Владислав Анатолійович

УДК 519.21

ДИСЕРТАЦІЯ

Дослідження асимптотичних властивостей випадкових дерев

113 — Прикладна математика

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Науковий керівник: д. ф.-м. н., проф. **Маринич Олександр Віталійович**

Київ – 2022

АНОТАЦІЯ

Богун В. А. Дослідження асимптотичних властивостей випадкових дерев.
– Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 113 – «Прикладна математика». – Київський національний університет імені Тараса Шевченка Міністерства освіти і науки України, Київ, 2022.

Дисертаційна робота присвячена дослідженню асимптотичних властивостей випадкових дерев, коли кількість вершин у дереві прямує до нескінченності. В роботі розглянуто три класи випадкових дерев, представлено результати про асимптотичну поведінку профілів, моди, ширини, висоти, довжин певних шляхів та інших характеристик. Першим класом є випадкові дерева простої будови, що є класичною моделлю в літературі. Її можна описати наступним чином: на кожному кроці в дереві навмання вибирається вершина з певної підмножини та до неї додаються нові вершини. До цього класу належать випадкові рекурсивні дерева та бінарні дерева пошуку. Другою моделлю слугують дерева загальних гіллястих процесів. Третіми виступають VP-дерева, що виникають в комп'ютерних науках як структура даних для швидкого виконання запиту пошуку найближчого сусіда.

В даній дисертаційній роботі продовжено дослідження випадкових дерев методами та техніками, використаними в нещодавніх роботах О. Іксанова, З. Каблучка, О. Маринича, Г. Зульцбаха та інших. При цьому отримано нові результати та збагачено загальну теорію, що використовується для доведень. Окрім цього в роботі досліджується новий клас дерев, що раніше не вивчався з ймовірнісної точки зору.

Автором дисертації отримано повні асимптотичні розклади майже напевне для профілів випадкових дерев простої будови, доведено граничні теореми для моди і ширини та запропоновано нові характеристики для випадкових величин, що при цьому виникають, у вигляді стохастичних рівнянь нерухомої точки. Проаналізовано висоту дерев загального гіллястого процесу побудованого за збуреним випадковим блуканням. Додатково автором отримано асим-

птотики згорток Лебега-Стільєса функцій лінійного росту, що з'являються при дослідженні подібних дерев на проміжних рівнях, а відповідні результати застосовані до функцій відновлення. Також було доведено слабкий закон великих чисел для довжини найлівішого шляху у VP-дереві та граничну теорему щодо її збіжності вздовж підпоследовностей. Нарешті, були дослідженні розбиття простору, що породжує випадкове VP-дерево. Отримано результат про скінченність моменту першого повернення в певні стани цих розбиттів, доведено існування стаціонарного розподілу відповідного ланцюга Маркова на множині розбиттів та збіжність до нього, а сам стаціонарний розподіл підраховано у явному вигляді .

Основні теоретичні результати супроводжуються математичним моделюванням та комп'ютерними симуляціями виконаними за допомогою мови програмування Python3, що підтверджують відповідні твердження та дають краще уявлення про досліджувані величини.

Дисертаційна робота має як теоретичне так і прикладне значення. Випадкові дерева мають численні застосування в аналізі структур даних та алгоритмів, а також для моделювання процесів еволюції та поширення в багатьох прикладних науках, зокрема біології, медицині, соціології, економіці та інших. Отримані в роботі результати можуть знайти застосування в згаданих сферах, а запропоновані математичні методи можна використовувати для подальших досліджень інших класів дерев.

Ключові слова: асимптотичний розклад, випадкові дерева, граничні закони, дерево загального гіллястого процесу, однорозгалуджене гіллясте випадкове блукання, пошук найближчого сусіда, випадкове дерево простої будови, рівняння нерухомої точки, VP-дерево.

SUMMARY

Bohun V. A. Asymptotic properties of random trees. – Manuscript.

Dissertation for scientific level of Doctor of Philosophy in speciality 113 – “Applied mathematics”. – Taras Shevchenko National University of Kyiv, Ministry of Education and Science of Ukraine, Kyiv, 2022.

The thesis is devoted to the analysis of asymptotic properties of random trees when the number of vertices goes to infinity. The work considers three classes of random trees and presents results on the asymptotic behaviour of profiles, mode, width, height, length of particular paths etc. The first class is simply generated random trees, which is a classic model in the literature. It can be described in the following way: on each step we pick a vertex uniformly at random from a subset of all vertices and add new vertices to the chosen one according to some prescribed rule. This class contains, in particular, random recursive trees and binary search trees. The second model is general branching process trees. Thirdly, we consider VP-trees, which appeared for the first time in computer science as a data structure for fast execution of nearest neighbor search queries.

In the present work we continue the analysis of random trees using methods and techniques developed in recent works of A. Iksanov, Z. Kabluchko, A. Marynych, H. Sulzbach et al. Not only we obtain new results but also improve general theory which is used in the proofs. Moreover, we analyse a new type of trees, which has not been studied before from a probabilistic viewpoint.

We obtain complete asymptotic expansions almost surely for profiles of simply generated random trees and provide limit theorems for mode and width, as well as propose new characterizations of random variables, which appear as the coefficients, via stochastic fixed-point equations. We also analyse the height of a general branching process tree of an iterated perturbed random walk. As a byproduct we derive new asymptotics for Lebesgue–Stieltjes convolutions of functions of linear growth which appear in the analysis of intermediate levels of trees of iterated perturbed random walk and apply these results to renewal functions. Besides we prove a weak law of large numbers for the length of the leftmost path in a VP-tree

and a limit theorem for this length along subsequences. Finally, we study partitions of a space generated by a random VP-tree. We prove the finiteness of the moment of the first return to certain states of these partitions, the existence of a stationary distribution of the corresponding Markov chain on the set of partitions and convergence this stationary distribution, which is also calculated explicitly.

Most of theoretical results are accompanied by mathematical modelling and computer simulations using a programming language Python3, which confirms the obtained assertions and gives a better understanding of studied quantities.

The thesis is both of theoretical and applied nature. Random trees have numerous applications in data structures and algorithms analysis, and in modelling of evolution and distribution processes in many applied sciences, in particular, in biology, medicine, sociology, economics etc. Obtained results can be used in the aforementioned areas, and developed mathematical methods can be used for future research of other classes of trees.

Keywords: asymptotic expansion, fixed-point equation, general branching process tree, limit theorems, nearest neighbour search, one-split branching random walk, random trees, simply generated random tree, VP-tree.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Публікації, в яких опубліковані основні наукові результати дисертації:

1. Bogun V. Almost sure asymptotic expansions for profiles of simply generated random trees // *Theory of Stochastic Processes*. – 2019. – **24(40)**, no. 1. – pp. 49-63.
2. Bohun V., Iksanov A., Marynych A., Rasytsov B. Renewal theory for iterated perturbed random walks on a general branching process tree: intermediate generations // *Journal of Applied Probability*. – 2022. – **59**, no. 2. – pp. 421-446.
3. Bohun V. Probabilistic analysis of vantage point trees // *Modern Stochastics: Theory and Applications*. – 2021. – **8(4)**. – pp. 413-434.
4. Bohun V., Marynych A. On the local time of a recurrent random walk on \mathbb{Z}^2 // *Theory of Probability and Mathematical Statistics*. – 2021. – **105**. – pp. 69-78.

Публікації, які засвідчують апробацію матеріалів дисертації:

1. Богун В. Асимптотичні розклади майже напевно для профілів випадкових дерев // VI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Об'єднані наукою: перспективи міждисциплінарних досліджень». – 2019. – сс. 85-87.
2. Bohun V. Simply generated random trees and their applications in computer science // *XDU-TSNUK Workshop on Probability Theory and Related Topics*. – 2020. – p. 4.
3. Богун В., Маринич О. Асимптотика згорток Лебега-Стілтєса функцій лінійного росту // Дев'ята міжнародна науково-практична конференція «Математика в сучасному технічному університеті» при КПІ ім. Ігоря Сікорського. – 2020. – сс. 37-39.

4. Bohun V. Probabilistic analysis of vantage point trees // 5th international conference «Modern Stochastics: Theory and Applications». – 2021. – p. 19.

ЗМІСТ

СПИСОК ПОЗНАЧЕНЬ	9
ВСТУП	11
Огляд літератури	26
Огляд за розділом 1	26
Огляд за розділом 2	32
Огляд за розділом 3	35
1 Випадкові дерева простої будови	38
1.1 Означення та обговорення	38
1.1.1 Однорозгалуджене гіллясте випадкове блукання.	38
1.1.2 Моделі випадкових дерев.	40
1.1.3 Функціонали на випадкових деревах.	42
1.2 Результати для випадкових дерев	43
1.2.1 Результати для зовнішніх профілів випадкових дерев.	48
1.2.2 Результати щодо моди випадкових дерев.	53
1.2.3 Результати щодо ширини випадкових дерев.	54
1.3 Математичне моделювання	56
1.3.1 Програмний код мовою Python3.	59
1.4 Висновки до розділу 1	66
2 Древа загальних гіллястих процесів	68
2.1 Означення та обговорення	68
2.2 Результати для кратних згорток функцій лінійного росту	71
2.3 Висота дерева загального гіллястого процесу	81
2.4 Математичне моделювання	82
2.4.1 Програмний код мовою Python3.	86

2.5	Висновки до розділу 2	90
3	VR-дерева	92
3.1	Означення та обговорення	92
3.2	Збіжність довжини найлівішого шляху	95
3.2.1	Аналіз процесу (X_h)	95
3.2.2	Граничні теореми для довжини найлівішого шляху. . .	100
3.3	Одновимірна збіжність	108
3.4	Балансування	117
3.5	Математичне моделювання	119
3.5.1	Програмний код мовою Python3.	124
3.6	Висновки до розділу 3	131
	ВИСНОВКИ	133
	Додаток	135
	Результати для однорозгалуджених гіллястих випадкових блукань .	135
	Додаткові твердження	137
	СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	137

СПИСОК ПОЗНАЧЕНЬ

\square – завершення доведення

\mathbb{R}_+ – невід’ємна півпряма $[0, \infty)$

\mathbb{N} – множина натуральних чисел $\{1, 2, \dots\}$

$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

\mathbb{Z} – множина цілих чисел

\mathbb{C} – множина комплексних чисел

$\overline{a, b} = \mathbb{Z} \cap [a, b]$

\Leftrightarrow – тоді і тільки тоді

в.в. – випадкова величина

м.н. – майже напевно

\xrightarrow{d} – збіжність за розподілом в.в.

$\xrightarrow{\mathbb{P}}$ – збіжність за ймовірністю в.в.

$\xrightarrow{\text{м.н.}}$ – збіжність майже напевно в.в.

$\stackrel{d}{=}$ – рівність розподілів випадкових елементів

$\stackrel{st}{\leq}$ – стохастичний порядок, $X \stackrel{st}{\leq} Y$ тоді і тільки тоді, коли $\mathbb{P}\{X \leq t\} \geq \mathbb{P}\{Y \leq t\}$ для всіх $t \in \mathbb{R}$

$f(x) = O(g(x))$ при $x \rightarrow x_0$, де $x_0 \in [-\infty, \infty]$, означає $\overline{\lim}_{x \rightarrow x_0} \frac{f(x)}{g(x)} < \infty$

$f(x) = o(g(x))$ при $x \rightarrow x_0$, де $x_0 \in [-\infty, \infty]$, означає $\lim_{x \rightarrow x_0} \left| \frac{f(x)}{g(x)} \right| = 0$

$f(x) \sim g(x)$ для функцій при $x \rightarrow x_0$, де $x_0 \in [-\infty, \infty]$, означає $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 1$

$X \sim Y$ для в.в. X та Y означає, що вони однаково розподілені

$\text{Geom}(p)$ – геометрично розподілена в.в. на \mathbb{N} (кількість випробувань) з параметром $p \in (0, 1]$

$\text{Exp}(\lambda)$ – експоненційно розподілена в.в. з параметром (інтенсивності) $\lambda > 0$

$\mathbb{1}_A$ – індикатор події A , що дорівнює 1, якщо подія A відбувається, та $= 0$, інакше

$x_+ = \max(x, 0)$

$\lfloor x \rfloor$ – ціла частина числа x

$\lceil x \rceil$ – округлення числа x до цілого вгору

$\text{frac}(x)$ – дробова частина числа x

$f * g$ – згортка Лебега-Стільтєса двох функцій локально обмеженої варіації

$f^{*(j)}$ – кратна згортка Лебега-Стільтєса функції f з собою j раз

δ_x – ймовірнісний розподіл, зосереджений в точці x

$\mathcal{L}(X)$ – функція розподілу в.в. X

\log – натуральний логарифм

\exp – експонента

\arg та Arg – аргумент та головний аргумент комплексного числа відповідно

ℓ_∞ – функція відстані максимуму двох векторів, тобто

$\ell_\infty((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1 \leq i \leq n} |x_i - y_i|$

ГВБ – гіллясте випадкове блукання

БДП – бінарне дерево пошуку

ВРД – випадкове рекурсивне дерево

ПОРД – планарно-орієнтоване рекурсивне дерево

ВСТУП

Обґрунтування вибору теми дослідження. Одним із актуальних напрямків сучасних досліджень в прикладній математиці, а точніше в прикладній теорії ймовірностей, є асимптотичний аналіз об'єктів та систем стохастичної природи. Об'єктом дослідження в даній дисертаційній роботі є випадкові дерева та пов'язані з ними моделі. Вони виникають при дослідженні майже будь-яких природних об'єктів, що мають ієрархічну структуру, при моделюванні різного роду еволюцій та розповсюджень, а також тісно пов'язані з великою кількістю структур даних та алгоритмів в комп'ютерних науках. Аналіз асимптотичних властивостей випадкових дерев полягає у доведенні граничних теорем різного характеру, і є предметом дослідження цієї роботи.

Випадкові дерева будуються, зазвичай, ітеративним шляхом, тобто з часом до дерева додаються нові елементи, а кожна ітерація містить в собі елемент випадковості. Інший підхід полягає у побудові дерева детермінованою процедурою з масиву вхідних даних, що припускається випадковим. В даній роботі розглядаються три класи випадкових дерев: випадкові дерева простої будови, дерева загальних гіллястих процесів та VP-дерева.

Дослідження за допомогою випадкових дерев проводились ще в XIX сторіччі. Класичним прикладом є дослідження генеологій за допомогою дерев процесів Гальтона-Ватсона. Найбільшої популярності та розвитку ця тематика набуває у другій половині XX сторіччя з розвитком комп'ютерних наук через їх велику прикладну цінність (див., наприклад, класичні роботи Д. Кнута, зокрема книгу [47]). З плином часу методика досліджень розвивалась та змінювалась на іншу, а відомі результати посилювались, як і з'являлись нові. Чергова хвиля досліджень виникає на початку XXI сторіччя, коли для доведень стали використовувати більш сучасні техніки, див., наприклад, роботи М. Дрмота, С.-К. Хвана, Л. Девройе, Х. Махмуда [15, 23, 25, 26, 54, 75], де найпопулярнішими підходами можна назвати використання теорії мар-

тингалів та метод сингулярного аналізу твірних функцій. Нещодавні роботи починають поєднувати дослідження дерев та аналіз інших об'єктів таких як гіллясті випадкові блукання. Так в роботі [43] 2017го року З. Каблучко, О. Мариничем та Г. Зульцбахом отримані результати для однорозгалуджених гіллястих випадкових блукань за допомогою техніки мод- ϕ збіжності, що була розроблена лише в останньому десятиріччі (див., наприклад, [39]). В роботах О. Іксанова, З. Каблучка, П. Ягерса, Л. Девройе, Б. Тота [22, 36, 37, 41, 67] загальні гіллясті блукання використовуються для аналізу випадкових, здебільшого рекурсивних, дерев. В даній дисертаційній роботі робиться акцент на використанні саме сучасних технік для аналізу випадкових дерев, а також їх подальшої розробки та вдосконалення. Додатково в роботі присутній аналіз класу дерев вперше представлений П. Янілосом у 1993 році (див. [79]), що до цього не вивчався з ймовірнісної точки зору.

Прикладне значення випадкових дерев дуже важко переоцінити. Вони знаходять застосування в економіці, соціальних дослідженнях, біології та медицині, зокрема, використовуються для моделювання розповсюдження вірусів. Дерева є невід'ємною складовою комп'ютерних наук, від базової роботи з даними та сортування до машинного навчання, а відповідними результатами вже класичних досліджень ми користуємось кожен день, працюючи з комп'ютером чи телефоном та навіть не звертаючи на це уваги. Окрім їх безпосередніх застосувань, дерева використовуються для отримання нових результатів в суміжних дисциплінах, наприклад, в аналізі алгоритмів.

Зв'язок роботи з науковими програмами, планами, темами. Перший розділ дисертаційної роботи виконано у відповідності до плану наукових досліджень кафедри дослідження операцій факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка в межах держбюджетної теми “Асимптотичний та структурний аналіз стохастичних моделей динаміки популяцій” (2019-2020 рр.), НДР №19БФ015-01. Також здобувач є учасником проєкту “Асимптотичні режими збурених випадкових блукань: на межі сучасної та класичної теорії ймовірностей” (2020-

2022 рр.) за підтримки Національного Фонду Досліджень України, проект 2020.02/0014, в планах наукових досліджень котрого виконано другий розділ дисертаційної роботи. Тематика досліджень дисертаційної роботи відповідає спеціальності «113 - Прикладна математика» галузі знань «Математика та статистика».

Мета і завдання дослідження. Початковою метою дисертаційної роботи є продовження досліджень асимптотичних властивостей випадкових дерев, що базуються на недавніх результатах, отриманих в вищезгаданих роботах Іксанова, Каблучка, Маринича та Зульцбаха. Проте надалі увага зосереджується на покращенні методів для майбутніх досліджень та аналізі нових класів дерев, що не розглядались раніше. Основним завданням автора було розширення області знань про випадкові дерева та прокладення шляху для нових досліджень в цьому напрямку. *Об'єкт дослідження* – випадкові дерева різних класів, а саме випадкові дерева простої будови, дерева загальних гіллястих процесів та VP-дерева, *предметом дослідження* є асимптотичні властивості цих об'єктів, тобто граничні властивості, коли кількість вершин в дереві прямує до нескінченності.

Методи дослідження. Дослідження асимптотичних властивостей випадкових дерев полягає в першу чергу в доведенні граничних теорем для різного роду характеристик (функціоналів) цих дерев чи пов'язаних з ними об'єктів. Функціонали, що фактично є випадковими величинами, в такому випадку збігаються після підхожого нормування та центрування до (можливо також випадкової) граничної величини в певному сенсі. Аналіз різних характеристик одного типу дерева, а особливо різних класів дерев, може повністю різнитись між собою, що призводить до використання великої кількості різних теорій, методів та технік. Зокрема в дисертаційній роботі використовуються такі методи прикладної математики та прикладної теорії ймовірностей як:

- метод стискаючих відображень та методи рівнянь нерухомої точки;
- теорія ймовірнісних метрик;

- теорія ланцюгів Макова та ланцюгів Гарріса;
- теорія відновлення та теорія загальних випадкових блукань;
- теорія гіллястих випадкових блукань;
- комбінаторні техніки та методи математичного аналізу;
- теорія характеристичних функцій;
- та інші.

Наукова новизна отриманих результатів. Наукова новизна цієї дисертаційної роботи полягає в розширенні та застосуванні відомих результатів для аналізу випадкових дерев простої будови, дослідження нових характеристик дерев загальних гіллястих процесів та випадкових VP-дерев.

Серед основних результатів дисертаційної роботи можна виділити наступні:

- отримані повні асимптотичні розклади майже напевне для профілів випадкових дерев простої будови;
- отримані граничні теореми для пов'язаних з профілями функціоналів, а саме ширини та моди;
- отримані нові характеристики для величин, що виникають у вищезгаданих розкладах;
- досліджено висоту дерева загального гіллястого процесу побудованого за збуреним випадковим блуканням;
- досліджено асимптотичні властивості згортки Лебега-Стільтєса для функцій лінійного росту у випадках, коли кількість елементів згортки розбігається до нескінченності;
- отримані результати застосовано для функцій відновлення, згортки котрих виникають у теорії ітерованих випадкових блукань та використовуються при дослідженні дерев;

- отримано збіжність підмножин, що породжуються випадковим VP-деревом, та підраховано сам граничний розподіл;
- отримано слабкий закон великих чисел для довжини найлівішого шляху у VP-дереві;
- отримана збіжність довжини найлівішого шляху VP-дерева вздовж під-послідовностей.

Теоретичне і практичне значення одержаних результатів. Результати дисертаційної роботи мають як теоретичний, так і прикладний характер. З теоретичної точки зору в роботі розширено теорію дерев загальних гіллястих процесів. Такі процеси використовуються, зокрема, для досліджень інших класів дерев, а отримані результати можуть бути корисними, наприклад, у вивченні проміжних рівнів профілів. Результати для VP-дерев можуть бути використані для подальших досліджень в цього класу дерев, а розроблені методи та техніки можуть слугувати відправною точкою для аналізу інших класів дерев, пов'язаних з задачею пошуку найближчого сусіда. З практичної точки зору велика кількість результатів, зокрема повні асимптотичні розклади та збіжності майже напевне для випадкових дерев простої будови, можуть в явному вигляді використовуватись у сферах застосувань дерев (див. розділ “Обґрунтування вибору теми дослідження”).

Особистий внесок здобувача. Визначення загального плану досліджень та постановка задач належить науковому керівникові. Всі результати, що виносяться на захист, належать здобувачеві.

У статті [12], написаній у співавторстві з Іксановим О. М., Мариничем О. В. та Рашитовим Б. С., Іксанову О. М. належить постановка задачі, Мариничу О. В. належить оформлення проміжних та фінальних версій, Рашитову Б. С. належать доведення твердження 3.3, лем 3.1 і 3.2 та теорем 2.1, 2.2 і 2.3, Богуну В. А. належать доведення тверджень 2.1, 3.1 і 3.2.

У статтях [10] та [11] Богуну В. А. належать всі представлені в них результати.

Апробація результатів дисертаційної роботи. Результати дисертаційної роботи доповідалися на:

- VI Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених “Об’єднані наукою: перспективи міждисциплінарних досліджень” (13-14 листопада 2019 року);
- міжнародному українсько-китайському воркшопі “Probability Theory and Related Topics” на базі університету Шидіан (Xidian) (22 грудня 2020 року);
- дев’ятій міжнародній науково-практичній конференції “Математика в сучасному технічному університеті” при КПІ ім. Ігоря Сікорського (28-29 грудня 2020 року);
- міжнародній конференції “Modern Stochastics: Theory and Applications V” (1-4 червня 2021 року).

Також матеріали дисертаційної роботи доповідались та обговорювались на наукових семінарах кафедри дослідження операцій Київського національного університету імені Тараса Шевченка.

Структура та обсяг дисертаційної роботи. Дисертаційна робота складається зі вступу, огляду літератури, трьох розділів та додатку. Кожен розділ розбито на підрозділи, які, в свою чергу, поділяються на пункти. В кінці кожного розділу містяться висновки до даного розділу. Розділи мають власну нумерацію формул, рисунків, таблиць та лістингів. Теореми, леми, твердження, зауваження, наслідки, приклади мають наскрізну нумерацію. Робота містить 17 рисунків, 3 таблиці та 12 лістингів. Список використаних джерел містить 79 позицій.

Допоміжні ресурси дисертаційної роботи. Протягом роботи для математичного моделювання та побудови графіків використовувалась мова програмування Python3, а в самій роботі наведений відповідний програмний код. Ця мова програмування та використані з нею ресурси, а саме включені до

неї бібліотеки та інтерпретатор мови, знаходяться під ліцензією PSF. Також в роботі використовуються додаткові бібліотеки третіх сторін, а саме NumPy (під ліцензією BSD-3), SciPy (під ліцензією BSD-3) та Matplotlib (під модифікованою ліцензією PSF). Усі перераховані засоби є ресурсами з відкритим вихідним кодом, а відповідні ліцензії дозволяють вільно їх використовувати для наукових досліджень.

Наведемо тепер перелік основних результатів дисертаційної роботи.

Випадкові дерева простої будови. Перший розділ присвячено однорозгалуженим гіллястим випадковим блуканням (ГВБ) та їх застосуванням для аналізу профілів випадкових дерев. Формально однорозгалужене ГВБ визначено наступним чином. Розглядається система частинок, що живуть на решітці \mathbb{Z} . В нульовий момент часу присутня лише одна частинка, розташована в 0. В кожний наступний дискретний момент часу обирається одна з живих частинок навмання та замінюється на кластер частинок, чії відносні зміщення відносно обраної частинки задаються точковим процесом $\zeta = \sum_{i=1}^N \delta_{Z_i}$. Тут випадкова величина N позначає кількість нових частинок, що утворюється. Ця величина вважається майже напевно скінченною.

Нехай у момент часу $n \in \mathbb{N}$ в системі знаходиться S_n частинок, а їх положення позначені як $x_{1,n}, \dots, x_{S_n,n}$. Через $m(\beta)$ позначимо твірну функцію моментів інтенсивності процесу ζ мінус 1:

$$m(\beta) := \mathbb{E} \left[\sum_{i=1}^N e^{\beta Z_i} \right] - 1.$$

Тоді нормалізована твірна функція моментів профілю визначається наступним чином:

$$W_n(\beta) := \frac{1}{n^{m(\beta)/m(0)}} \sum_{i=1}^{S_n} e^{\beta x_{i,n}}.$$

За виконання п'яти простих умов для процесу ζ , що перелічені в лемі 1, всі ці функції коректно визначені в деякому околі нуля комплексної площини. Більш того, W_n збігається майже напевне, при $n \rightarrow \infty$, до випадкової аналітичної функції W_∞ в цьому околі.

Виявляється, що випадкова величина $W_\infty(\beta)$ задовольняє рівняння нерухомої точки.

Теорема 2. Для однорозгалудженого ГВБ такого, що твердження лемми 1, яка наведена в основній частині роботи, виконуються та точковий процес для зміщень ζ детермінований й складається з N точок, випадкова величина $W_\infty(\beta)$ задовольняє наступне стохастичне рівняння нерухомої точки

$$W_\infty(\beta) \stackrel{d}{=} \sum_{i=1}^N e^{\beta Z_i} U_i^{m(\beta)/m(0)} W_\infty^{(i)}(\beta),$$

де $W_\infty^{(i)}(\beta)$ незалежні в сукупності копії $W_\infty(\beta)$, що не залежать від випадкового вектора $(U_i)_{i=1}^N$, який має розподіл Діріхле з параметром $\frac{1}{N-1}$.

Позначимо логарифмічні похідні функції $W_\infty(\beta)$ через $\chi_j(\beta) = (\log W_\infty)^{(j)}(\beta)$ ($j \in \mathbb{N}$). Ця теорема також дозволяє знаходити й рівняння нерухомої точки для величин $\chi_j(0)$.

Наведемо тепер класичну формальну модель випадкового дерева простої будови. Поділимо всі вершини дерева на зовнішні та внутрішні. В нульовий момент часу дерево складається лише з його кореня, що є зовнішньою вершиною. В кожний наступний дискретний момент часу навмання обирається одна зовнішня вершина, вона зазвичай стає внутрішньою та до неї приєднуються нові, можливо випадкові, піддерева, що містять внутрішні та зовнішні вершини.

Розглянемо стан дерева на n -му кроці побудови. Рівнем вершини називатимемо її відстань від кореня. Позначатимемо через $\mathbb{L}_n(k)$ (зовнішній) профіль дерева, тобто кількість зовнішніх вершин дерева на рівні k . Шириною дерева M_n називатимемо максимальну кількість зовнішніх вершин, що знаходяться на одному рівні, а модою u_n називатимемо рівень, де цей максимум досягається.

Випадкові дерева простої будови природним чином описуються однорозгалудженими ГВБ, як пояснюється в розділі 1. В ньому розглянуто профілі випадкових рекурсивних дерев, D -арних дерев та p -орієнтованих рекурсив-

них дерев. В D -арних деревах зовнішніми вершинами позначаються листки, а на кожній ітерації до рівномірно обраного листка додаються D синів. На прикладі цих дерев мають місце наступні твердження.

Теорема 7. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D-1)n+1$ зовнішніми вершинами. Для всіх $r \in \mathbb{N}_0$ маємо

$$(\log n)^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \left| \frac{1}{n} \mathbb{L}_n(k) - H_n(k) \sum_{j=0}^r G_j \left(\frac{k - \frac{D}{D-1} \log n}{\sqrt{\frac{D}{D-1} \log n}}; 0 \right) \frac{1}{(\log n)^{j/2}} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де поліном G_j наведено в зауваженні 10 в основній частині роботи, та

$$H_n(k) = \frac{(D-1)^{3/2}}{\sqrt{2D\pi \log n}} e^{-\frac{(k - \frac{D}{D-1} \log n)^2}{\frac{2D}{D-1} \log n}}.$$

Теорема 11. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D-1)n+1$ зовнішніми вершинами. Для всіх $r \in \mathbb{N}_0$ та кожної компакної множини $L \subset (\frac{D}{D-1}e^{\beta_-}, \frac{D}{D-1}e^{\beta_+})$ виконується

$$(\log n)^{r+1} \sup_{k \in \mathbb{Z} \cap (\log n)L} \left| H_n(k) \mathbb{L}_n(k) - \frac{1}{\sqrt{2\pi k}} \sum_{j=0}^r \frac{F_{2j}(0; \log \frac{(D-1)k}{D \log n})}{(\log n)^j} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де $F_{2j}(0; \beta) = W_\infty(\beta) G_{2j}(0; \beta)$, константи β_- та β_+ є розв'язками $De^\beta(1-\beta) = 1$ й

$$H_n(k) = n^{\frac{1}{D-1}} \left(\frac{(D-1)k}{De \log n} \right)^k.$$

Теорема 15. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D-1)n+1$ зовнішньою вершиною. Покладемо $\mathbb{L}_n^\circ(k) := \mathbb{L}_n(k) - \mathbb{E}[\mathbb{L}_n(k)]$ та нехай $(k_n)_{n \in \mathbb{N}}$ позначає детерміновану цілочисельну послідовність.

(а) Якщо $k_n = \frac{D}{D-1} \log n + \alpha \sqrt{\frac{D}{D-1} \log n} + o(\sqrt{\log n})$ для деякого $\alpha \in \mathbb{R}$, тоді

$$\frac{\log n}{n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)^2 \alpha e^{-\frac{1}{2}\alpha^2}}{D\sqrt{2\pi}} (\chi_1(0) - \mathbb{E}[\chi_1(0)]).$$

(б) Якщо $k_n = \frac{D}{D-1} \log n + c_n$, де $c_n = o(\log n)$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n e^{c_n}} \left(\frac{(D-1)k_n}{D \log n} \right)^{k_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)(\chi_1(0) - \mathbb{E}[\chi_1(0)])}{(\frac{D}{D-1})^{3/2} \sqrt{2\pi}}.$$

Зокрема, якщо $c_n = o(\sqrt{\log n})$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)(\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\left(\frac{D}{D-1}\right)^{3/2} \sqrt{2\pi}}.$$

(в) Якщо $k_n = \frac{D}{D-1} \log n + c_n$, де $c_n = O(1)$, тоді

$$\frac{(\log n)^{3/2}}{n} \mathbb{L}_n^\circ(k_n) - \frac{D-1}{\left(\frac{D}{D-1}\right)^{3/2} \sqrt{2\pi}} (R(c_n) - \mathbb{E}[R(c_n)]) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де $R(c_n) = (c_n + \frac{1}{2})\chi_1(0) - \frac{\chi_1^2(0) - \chi_2(0)}{2}$.

З використанням результатів для профілів отримано результати для ширини та моди.

Теорема 19. Існує м.н. скінченна випадкова величина K така, що для всіх $n > K$ мода u_n для D -арного дерева з $(D-1)n + 1$ зовнішньою вершиною рівна одному з чисел

$$\left\lfloor \frac{D}{D-1} \log n + \chi_1(0) - \frac{1}{2} \right\rfloor \text{ або } \left\lceil \frac{D}{D-1} \log n + \chi_1(0) - \frac{1}{2} \right\rceil.$$

Теорема 23. Нехай M_n – ширина D -арного дерева з $(D-1)n + 1$ зовнішньою вершиною, тоді виконується

$$\frac{\sqrt{2D\pi \log n}}{(D-1)^{3/2} n} M_n \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1.$$

Теорема 26. Нехай M_n – ширина D -арного дерева з $(D-1)n + 1$ зовнішньою вершиною. З ймовірністю 1 множина границь підпоследовностей послідовності

$$\tilde{M}_n := 2 \frac{D}{D-1} \log n \left(1 - \frac{\sqrt{2\pi \log n} \sqrt{\frac{D}{D-1}} M_n}{(D-1)n} \right), \quad n \in \mathbb{N},$$

є інтервалом $[\chi_2(0) - 1/12, \chi_2(0) + 1/6]$. Більш того, якщо покласти $\theta_n := \min_{k \in \mathbb{Z}} \left| \frac{D}{D-1} \log n + \chi_1(0) - 1/2 - k \right|$, то виконується

$$\tilde{M}_n - \theta_n^2 \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \chi_2(0) - \frac{1}{12}.$$

Деякі з цих результатів є відомими в літературі у випадку $D = 2$, або в менш загальному вигляді у сенсі, коли присутні лише перші члени асимптотичного розкладу.

Дерева загальних гіллястих процесів. Другий розділ дисертаційної роботи присвячено деревам загальних гіллястих процесів. Під час їх дослідження виникають згортки функцій лінійного росту, для котрих отримано наступні загальні твердження.

Згортку Лебега-Стільтєса двох функцій f та g локально-обмеженої варіації будемо позначати наступним чином

$$(f * g)(t) = \int_{\mathbb{R}} f(t - y) dg(y), \quad t \in \mathbb{R}.$$

Згортку функції f з собою j раз позначимо через $f^{*(j)}$.

Теорема 29. *Нехай $f : \mathbb{R} \rightarrow [0, \infty)$ позначає неспадну неперервну справа функцію, що обертається в нуль на від'ємній півпрямій, та задовольняє*

$$f(t) = at + O(t^\alpha), \quad t \rightarrow \infty$$

для деяких $a > 0$ та $\alpha \in [0, 1)$. Тоді, для будь-якої цілозначної функції $j = j(t)$ такої, що $j(t) = o(t^{(1-\alpha)/2})$ при $t \rightarrow \infty$, виконується

$$f^{*(j)}(t) \sim \frac{a^j t^j}{j!}, \quad t \rightarrow \infty.$$

Теорема 31. *Нехай $f : \mathbb{R} \mapsto [0, \infty)$ позначає неспадну неперервну справа функцію, що обертається в нуль на від'ємній півпрямій. Припустимо, що функція ε , визначена рівністю*

$$\varepsilon(t) := f(t) - at, \quad t \geq 0,$$

для деякого $a > 0$, задовольняє

$$\int_{[0, \infty)} y d(\mathcal{V}_{[0, y]}(\varepsilon)) < \infty,$$

де $\mathcal{V}_{[0, y]}(\varepsilon)$ позначає повну варіацію ε на інтервалі $[0, y]$.

Тоді, для будь-якої цілочисельної функції $j = j(t)$ такої, що $j(t) = o(t^{2/3})$ при $t \rightarrow \infty$, виконується

$$f^{*(j)}(t) \sim \frac{a^j t^j}{j!} \exp\left(\frac{\gamma_0 j^2}{at}\right), \quad t \rightarrow \infty,$$

де $\gamma_0 := \int_{[0, \infty)} d\varepsilon(y) = \lim_{t \rightarrow \infty} (f(t) - at)$.

Відповідні твердження застосовані для функцій відновлення, згортки котрих виникають при дослідженні ітерованих випадкових блукань та їх збурених версій.

Наведемо схему побудови загального гіллястого процесу згенерованого за збуреним випадковим блуканням. Нехай $(\xi_i, \eta_i)_{i \in \mathbb{N}}$ – послідовність незалежних копій випадкового вектора (ξ, η) з м.н. додатними компонентами. Через $S_0 = 0$ та $S_i = \sum_{k=1}^i \xi_k$ ($i \in \mathbb{N}$) позначимо стандартне випадкове блукання. Покладемо $T_i := S_{i-1} + \eta_i$ ($i \in \mathbb{N}$), тоді послідовність $T := (T_i)_{i \in \mathbb{N}}$ називається збуреним випадковим блуканням.

В момент часу 0 присутня єдина особа. Вона продукує потомство (перше покоління), часи народження котрого задані точками T . На кожному кроці кожна особа в поточному поколінні продукує нащадків в наступному, а їх часи народження відносно часу народження батька розподілені згідно копії T , й для різних батьків ці копії незалежні.

Для $t \geq 0$ та $j \in \mathbb{N}$ позначимо через $N_j(t)$ кількість осіб в j -му поколінні з часами народження не більшими за t . Тоді величина

$$H(t) = \inf\{j \in \mathbb{N} : N_j(t) = 0\}, \quad t > 0,$$

називається висотою дерева загального гіллястого процесу згенерованого за збуреним випадковим блуканням T та обмеженого смугою $[0, t]$. Для неї отримано наступну граничну теорему.

Теорема 34. Для всіх $t \geq 0$, $H(t) < \infty$ м.н. Більш того,

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{\gamma} \in (0, \infty) \quad \text{м.н.},$$

де $\gamma := \sup\{z > 0 : \mu(z) < 1\}$ та $\mu(z) := \inf_{s > 0} (e^{zs} \frac{\mathbb{E}e^{-s\eta}}{1 - \mathbb{E}e^{-s\xi}})$ при $z > 0$.

VP-дерева. Третій розділ дисертаційної роботи присвячено дослідженню випадкових VP-дерев з метрикою ℓ_∞ в просторі $[-1, 1]^d$ ($d \geq 1$). Нехай задана послідовність випадкових точок x_1, x_2, \dots , де елементи x_i незалежні та мають рівномірний розподіл в $[-1, 1]^d$. Випадкове VP-дерево є бінарним деревом, що будується за такою ітеративною схемою. Перша точка x_1 завжди зберігається в корені дерева, також збережемо в цій вершині величину $\tau \in (0, 1)$, що називатимемо пороговим значенням. Наступні точки додаються послідовно за допомогою рекурсивної процедури, що починається від кореня. Припустимо, що точка x_n додається до вершини T . Якщо ця вершина порожня, то в ній зберігаються ця точка та порогове значення τ^{h_T} , де h_T позначає відстань від T до кореня плюс один, й процедура додавання завершується. Якщо ж вершина T не порожня, то в ній збережені деяка точка x_T та порогове значення τ^{h_T} . В цьому випадку якщо $\ell_\infty(x_T, x_n) \leq \tau^{h_T}$, то процедура продовжується рекурсивно з лівого сина T , а інакше – з правого.

Найлівішим шляхом довільного бінарного дерева називатимемо шлях, що починається в його корені та завжди прямує до лівого сина поточної вершини поки не досягне листка. Через I_h позначимо підмножину простору $[-1, 1]^d$, що містить всі точки x такі, що при додаванні в дерево вони були б збережені в лівому піддереві T_h , де послідовність $(T_h)_{h \in \mathbb{N}}$ позначає вершини дерева, що утворюють найлівіший шлях у ньому починаючи з кореня, та $I_0 = [-1, 1]^d$. Через $X_h := I_h / \tau^h$ ($h \in \mathbb{N}_0$) позначатимемо нормовані розбиття простору. Наступні твердження стосуються основних результатів щодо цієї послідовності.

Теорема 39. *Нехай α позначає будь-яке можливе нормоване розбиття простору, що породжене найлівішим шляхом випадкового VP-дерева, та*

$$R_\alpha := \inf\{h \geq 1 : X_h = [-1, 1]^d \text{ за умови, що } X_0 = \alpha\}.$$

Тоді $\mathbb{E}R_\alpha < \infty$.

Теорема 47. *Послідовність $(X_h)_{h \geq 0}$ є ланцюгом Маркова та має стаціонар-*

ний розподіл I_∞ й

$$\lim_{n \rightarrow \infty} \|\mathbb{P}\{X_h = \cdot | X_0 = [-1, 1]^d\} - I_\infty(\cdot)\|_{TV} = 0,$$

де $\|\cdot\|_{TV}$ позначає відстань повної варіації.

Наступний результат дозволяє явно знайти одновимірні складові стаціонарного розподілу, позначні через X_∞ , у випадку, коли точки послідовності для побудови дерева розподілені рівномірно. Якщо X є випадковою величиною будемо позначати через $\mathcal{L}(X)$ її розподіл.

Теорема 53. Відображення $\mathcal{L}(X) \mapsto \mathcal{L}(f(X))$ в просторі ймовірнісних мір на $[1, 2]$, що наділений ℓ_1 -мінімальною метрикою, де

$$f(X) = \min\left\{\frac{uX}{\tau}, 1\right\} + \min\left\{\frac{(1-u)X}{\tau}, 1\right\}$$

є строго стискаючим, де u рівномірно розподілена випадкова величина на $[0, 1]$, незалежна від випадкової величини X .

Більш того в роботі наведена техніка для підрахунку нерухомої точки X_∞ стохастичного рівняння

$$X_\infty \stackrel{d}{=} \min\{\tau^{-1}uX_\infty, 1\} + \min\{\tau^{-1}(1-u)X_\infty, 1\}.$$

Його явний розв'язок сильно залежить від значення параметра $\tau \in (0, 1)$. Так, наприклад, коли $\tau < \frac{1}{2}$, то $X_\infty = p(1+u) + (1-p)\delta_2$, де

$$p = \frac{1}{1 - 2 \log 2 + \frac{1}{\tau}}.$$

Для розміру найлівішого шляху VP-дерева L_n , тобто кількості вершин у ньому на n -му кроці побудови дерева, доведено наступні теореми.

Теорема 48. Для випадкового VP-дерева розмір найлівішого шляху L_n задовольняє наступний слабкий закон великих чисел, при $n \rightarrow \infty$,

$$\frac{L_n}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{-\log \tau}.$$

Теорема 52. Для довільного фіксованого $T > 0$

$$L_{\lfloor T\tau^{-n} \rfloor} - n \xrightarrow{d} L_{\infty}^{(T)}$$

при $n \rightarrow \infty$ та функція розподілу $L_{\infty}^{(T)}$ рівна

$$\mathbb{P}\{L_{\infty}^{(T)} \leq m\} = 1 - \mathbb{P}\{G_{\infty} \leq T\tau^m\}$$

для всіх $m \in \mathbb{Z}$.

Тут G_{∞} позначає випадковий ряд $\sum_{j=1}^{\infty} \xi_j$, де $(\xi_j)_{j \geq 0}$ є умовно незалежними, при фіксованому I_{∞} , випадковими величинами такими, що ξ_j має показниковий розподіл з $\text{Exp}(\frac{|I_{\infty}|}{2^d \tau^j})$, де $|I_{\infty}|$ означає міру Лебега випадкової множини I_{∞} .

Подяка. Автор дисертації Богун В. А. висловлює щирю вдячність своєму науковому керівникові – доктору фізико-математичних наук, професору Мариничу Олександрові Віталійовичу – за плідні дискусії й обговорення наукових проблем, за влучні поради при оформленні статей та дисертаційної роботи, а понад усе, за справжню турботу і постійну увагу протягом усього часу написання роботи, без яких вона навряд чи була б зроблена. Також автор висловлює подяку всім колегам з кафедри дослідження операцій факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка за їх підтримку та можливість спільно працювати в рамках проєкту Національного Фонду Досліджень України. Окрему подяку за допомогу в подоланні спільних труднощів автор висловлює Рашитову Богдану Сергійовичу.

Огляд літератури

Огляд за розділом 1

В цьому розділі розглядаються асимптотичні властивості випадкових дерев простої будови. Такі дерева побудовано за наступною рекурсивною процедурою. На початку присутнє певне дерево (зазвичай, лише корінь). Всі вершини дерева умовно поділено на внутрішні та зовнішні. На кожному кроці рівномірно обирається випадкова зовнішня вершина та замінюється на певне піддерево, що може містити нові внутрішні та зовнішні вершини. У книзі [24] можна знайти більш детальний огляд випадкових дерев.

Основною характеристикою, що досліджується в першому розділі, є зовнішній профіль дерева (або просто профіль) $(\mathbb{L}_n(k))_{k \in \mathbb{Z}}$, де $\mathbb{L}_n(k)$ – кількість зовнішніх вершин на рівні k на n -ій ітерації побудови дерева, де під рівнем вершини розуміється її відстань до кореня. Серед інших відомих характеристик дерев можна виділити

- висоту – максимальну довжину від кореня до іншої вершини;
- ширину – максимальне значення профілю дерева;
- моду – рівень, де досягається ширина;
- повну довжину – сума рівнів всіх зовнішніх вершин;
- та інші.

Дослідження випадкових дерев простої будови можна поділити на три категорії в залежності від типу дерева.

Найпростішою моделлю є випадкове рекурсивне дерево. В цій моделі всі вершини є зовнішніми та на кожній ітерації до будь-якої навантаженої вершини додається одна нова. Не зважаючи на просту побудову ця модель знаходить багато практичних застосувань, зокрема в моделюванні розповсюдження хвороб, розвитку мов в лінгвістиці, поведінки рекурсивних алгоритмів в комп'ютерних науках, пірамідальних схем в економіці, представленні локальних та інтернет мереж й процесів, що виникають в них, та багато іншого. В роботі [75] можна знайти більш детальний огляд їх застосувань. Деякі нещодавні результати щодо профілів випадкових рекурсивних дерев чи пов'язаних з ними характеристик можуть бути знайдені в роботах [23, 25, 32].

Найпопулярнішою моделлю випадкового дерева є бінарне дерево пошуку. Це дерево виникло як структура даних для швидкого пошуку елементів, й будується воно за наступною схемою. Нехай дана послідовність $(u_n)_{n \in \mathbb{N}}$ незалежних однаково розподілених випадкових величин з рівномірним на $[0, 1]$ розподілом¹. На першому кроці дерево складається лише з кореня, в якому зберігається елемент u_1 . Всі наступні кроки виконуються за наступною рекурсивною процедурою. Нехай дерево вже складається з $n - 1$ вершин, тоді при додаванні елемента u_n він спочатку порівнюється з елементом в корені. Якщо $u_n < u_1$, то вставка рекурсивно продовжується з лівого сина кореня, інакше з правого. Так процедура порівняння повторюється з поточним розглядуваним елементом в дереві й аналогічно продовжується, поки відповідного сина не існує. В такому випадку він додається, й елемент u_n зберігається в ньому. Таким чином бінарне дерево пошуку підпорядковується лише одному простому правилу: для будь-якої вершини дерева й елемента u , що збережений в ній, всі елементи в лівому піддереві менші за u , а всі елементи, що знаходяться у правому піддереві цієї вершини, більші за u . Простими комбінаторними підрахунками можна пересвідчитись, що ця модель еквівалентна схемі побудови випадкового дерева простої будови, де лиски дерева – це зовнішні вершини,

¹Класично розглядати саме рівномірний розподіл, проте u_n можуть мати будь-який неперервний розподіл. Важливим критерієм є умова, що елементи послідовності незалежні однаково розподілені та не повторюються з ймовірністю 1.

й на кожній ітерації до випадкового листка додаються 2 зовнішні вершини, а саме обраний листок стає внутрішньою вершиною. Бінарне дерево пошуку є одним із найбільш досліджених типів випадкових дерев, дивись, наприклад, [15, 16, 23, 25, 26, 32] для нещодавніх результатів щодо їх профілів.

Очевидним застосуванням цих дерев є дослідження характеристик вищенаведеної структури даних для швидкого пошуку елементів, що має велике прикладне значення. Так, наприклад, було доведено, що середній рівень навімання вибраної вершини в дереві рівний $2 \log n + O(1)$, а максимальний рівень (чи висота) веде себе як $\beta \log n$, де $\beta \approx 4.311$, див. [65]. Ці характеристики відповідають середньому часу пошуку елемента в дереві та верхній оцінці на час додавання нового елемента.

Іншим застосуванням випадкових бінарних дерев є їх зв'язок з рекурсивними алгоритмами типу “поділяй-та-володарюй”. Зокрема, найвідомішим прикладом є алгоритм швидкого сортування або алгоритм Хоара. Щоб провести аналогію достатньо розглянути список з n елементів $(u_i)_{i=1..n}$, що визначені як i в бінарному дереві пошуку, й на кожній ітерації обирати ведучим елементом найперший елемент поточного списку. Тоді якщо зобразити всі рекурсивні виклики алгоритму у вигляді дерева, а в кожній вершині зберегти ведучий елемент, що обирався, то це дерево буде в точності бінарним деревом пошуку, побудованим на послідовності $(u_i)_{i=1..n}$. Таким чином відомі результати для бінарних дерев переносяться й на алгоритм швидкого сортування, а самі результати стверджують високу оптимальність алгоритму. Хоча послідовності для сортування зазвичай не випадкові й елементи в них можуть повторюватись, достатньо застосувати декілька евристик (наприклад, обирання ведучого елемента навімання), щоб навіть ці послідовності ставали еквівалентними бінарним деревам пошуку. Поєднання цих простих евристик та класичних результатів для бінарних дерев гарантувало, що алгоритм Хоара є найпопулярнішим алгоритмом для сортування на сьогодні, див., наприклад, [47].

Нарешті, завершуючи з застосуваннями, зауважимо, що як i в випадку

з сортуванням, на практиці послідовність $(u_n)_{n \in \mathbb{N}}$ зазвичай не є випадковою. Якщо покласти $u_n = n$, то побудоване бінарне дерево виродиться в так зване лінійне дерево, а пошук елементів стане еквівалентним їх перебору. З ідеєю модифікувати побудову дерева так, щоб фінальна структура стала випадковою навіть коли послідовність не випадкова, було створене так зване рандомізоване бінарне дерево пошуку ([57]), що дозволяє нескладним чином оптимізувати структуру.

До третьої категорії випадкових дерев можна віднести решту типів дерев, й навіть сумарно вони матимуть меншу кількість досліджень ніж випадкові рекурсивні дерева чи бінарні дерева пошуку. Основними класами тут є якраз узагальнення цих дерев.

Узагальненням бінарних дерев пошуку є D -арні дерева, коли замість 2 синів додається D синів. Функціональні граничні теореми для профілів цих дерев можуть бути знайдені в статті [70]. Зазначимо також, що при $D > 2$ означення D -арного дерева не є тим самим, що й означення m -арного дерева пошуку. Це розширення є суто структурним, але не природнім. Природними узагальненням бінарних дерев пошуку є щойно названі m -арні дерева пошуку або дерева квадрантів, проте обидві ці моделі на є деревами простої будови. Для комплексного аналізу дерев пошуку ми посилаємось на книгу [24].

До узагальнень моделі випадкового рекурсивного дерева відносяться планарно-орієнтовані рекурсивні дерева. Фактично це рекурсивне дерево, де також враховується розташування чергової вершини (зліва чи справа відносно сусідніх). Ми посилаємось на [35, 44, 76] для деяких нещодавніх результатів щодо профілів цих дерев. Серед їх застосувань можна виділити ті самі що й для звичайних рекурсивних дерев. Зокрема, планарні дерева нерідко дозволяють краще моделювати інтернет мережі, соціальні мережі та розповсюдження інформації в них.

Серед основних методів та технік досліджень випадкових дерев можна виділити наступні:

- Дослідження рівнянь нерухомої точки та метод стискаючих відобра-

жень. Враховуючи рекурсивну структуру побудови випадкових дерев, дуже природно описувати їх характеристики через рекурентні формули, що в граничному випадку формально зводяться до рівнянь нерухомої точки. Тому не дивно, що ця методика є найрозповсюдженішою, а при дослідженні нових об'єктів цей підхід варто мати на увазі одним із перших.

- **Метод моментів та метод твірних функцій.** Це також один із перших підходів до досліджень дерев, в ньому потрібно знайти твірну функцію досліджуваної характеристики, що в подальшому дозволяє знайти граничну функцію й відновити властивості граничної характеристики. Зокрема інколи знаходять всі моменти випадкової величини, що дозволяє її повністю описати. Див., наприклад, [54] для однієї з останніх робіт, що застосовує цей підхід.
- **Методи теорії мартингалів.** В цьому підході за правильного зображення досліджувана характеристика представляє собою мартингал. Для прикладу, нормована твірна функція профілю деяких класів дерев є мартингалом. Середнє значення характеристики зазвичай вже відоме й отримане більш простими методами. Основною частиною є саме доведення, що дана характеристика є мартингалом, а потрібні результати потім випливають із вже відомих тверджень теорії мартингалів. Такий підхід є більш сучасним та більш сильним у тому розумінні, що дозволяє, наприклад, доводити збіжність у сенсі майже напевне, що не вдається зробити попередніми двома методами. Саме цим методом була отримана більшість результатів щодо профілів та ширини, посилання на які були наведені при огляді класів дерев.
- **Комбінаторний підхід, тобто явний підрахунок характеристик дерев за допомогою комбінаторних міркувань.** Суто теоретично цим методом можна встановити всі характеристики дерев, проте доведення занадто швидко стають громіздкими. Цей метод, зазвичай, застосовується лише

для доведення базових властивостей (див., наприклад, [58]), хоча потенційно може стати популярним з розвитком комп'ютерних доведень.

- Методи гіллястих випадкових процесів та теорії відновлення. Цей підхід не є новим (див., наприклад, [22]), проте в останні роки він набуває нової хвилі популярності, з розвитком суміжних теорій. Його ідея полягає в описі характеристик дерев через характеристики інших об'єктів, таких як випадкові блукання чи загальні гіллясті процеси. Для прикладів дивись [34, 36, 37, 38].

Можна побачити, що у більш сучасних методах та техніках доведень використовуються результати суміжних теорій. Це в свою чергу дозволяє отримувати більш повні результати, або доводити вже відомі твердження простішими способами. Так в нещодавній роботі Каблучка, Маринича та Зульцбаха ([43]) наводяться результати стосовно однорозгалужених гіллястих випадкових блукань. Головним інструментом в доведенні цих результатів було поняття мод- ϕ збіжності, або більш точно мод-Пуассон збіжності. Не вдаючись в усі технічні деталі, поняття мод-Пуассон збіжності можна описати наступним чином.

Нехай $(\xi_n)_{n \in \mathbb{N}}$ – послідовність випадкових величин зі значеннями на \mathbb{R} , чий перетворення Лапласа $\mathbb{E}e^{s\xi_n}$ є скінченними для всіх $s \in \mathbb{C}$, та $(\lambda_n)_{n \in \mathbb{N}}$ – послідовність додатних чисел, для якої $\lim_{n \rightarrow \infty} \lambda_n = +\infty$. Кажуть, що послідовність $(\xi_n)_{n \in \mathbb{N}}$ збігається в сенсі мод-Пуассона зі швидкістю $(\lambda_n)_{n \in \mathbb{N}}$, якщо збіжність

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}e^{s\xi_n}}{e^{\lambda_n(e^s - 1)}} = A(s)$$

є локально рівномірною в деякій відкритій множині $\mathcal{D} \subset \mathbb{C}$, що містить дійсну вісь. Тут $A : \mathcal{D} \rightarrow \mathbb{C}$ є деякою аналітичною функцією, а в роботі ми спостерігатимемо саме логарифмічну швидкість, тобто $\lambda_n = \log n$. Знаменник у виразі є твірною функцією випадкової величини з розподілом Пуассона з параметром λ_n , таким чином цю границю можна розуміти як наближення

вигляду

$$\xi_n \stackrel{d}{=} \mathcal{P}_n + \Xi + o(1),$$

де Ξ є “випадковою величиною” з твірною функцією моментів $\mathbb{E}e^{s\Xi} = A(s)$, що не залежить від \mathcal{P}_n , що має розподіл Пуассона з параметром λ_n , та $o(1)$ збігається до 0 за ймовірністю. Більш точні означення цього типу збіжності та деякі загальні результати можна знайти в [29], та побачити в [39] і [48] приклади її застосування.

Основна мета першого розділу дисертаційної роботи полягає у розв’язанні трьох задач. По-перше, ми отримуємо асимптотичні розклади для профілів, моди та ширини випадкових дерев простої будови, таких як D -арні дерева, випадкові рекурсивні дерева та p -орієнтовані рекурсивні дерева, коли число вершин в них прямує до нескінченності. Ця частина базується на результатах роботи [43], що були отримані для однорозгалуджених гіллястих випадкових блукань, й використовує порівняння цих характеристик з характеристиками випадкових дерев простої будови. Таким чином, доведено повні асимптотичні розклади м.н. для профілів, а отримані результати порівнюються з вже відомими в літературі окремими випадками. Друга мета цього розділу – навести нові характеристики випадкових коефіцієнтів, що виникають у вищеназваних результатах за допомогою виведення точних стохастичних рівнянь нерухомої точки для їх розподілів. Нарешті, третьою складовою мети є чисельне моделювання, що підтверджує отримані теоретичні результати.

Огляд за розділом 2

В цьому розділі розглядаються асимптотичні властивості дерев загальних гіллястих процесів. Гіллясті процеси тісно пов’язані з випадковими деревами, це легко побачити, оскільки еволюцію гіллястого процесу можна зображати випадковими деревами, а їх характеристики нерідко можна закодувати один через інший. Для загального огляду спільних властивостей дерев та гіллястих процесів ми посилаємось на [27, 34], а тут розглянемо лише ті, що пов’язані

з дисертаційною роботою.

Гіллясті процеси застосовуються в тих сферах, що й випадкові дерева простої будови, коли потрібно розглядати модель в неперервному часі. Наприклад, природно розглядати поширення інформації в інтернеті чи інфікування осіб в неперервному часі ([3, 7, 40, 46]), що як раз робиться розширюючи модель випадкового рекурсивного дерева, а найперші спільні дослідження дерев та гіллястих процесів виникли при аналізі еволюції прізвищ у населення.

Тут розглядатиметься лише частковий випадок загального гіллястого процесу (або процесу Крампа-Мода-Ягерса), що заданий наступним чином. Нехай дано деякий точковий процес Z . В нульовий момент часу присутня лише одна особа. В перший момент часу ця особа породжує перше покоління, особи котрого знаходяться в точках Z . Кожна особа першого покоління незалежно від інших продукує синів, що належать наступному поколінню, а їх розташування відносно батька розподілені як незалежна копія Z , і так далі. В такий процес можна вкласти еволюцію дискретного дерева, використовуючи правильні моменти зупинки. Хоча подібна ідея та методика не є новою (див. [22, 64]), вона набуває нової хвилі популярності саме зараз за рахунок розвитку в суміжних дисциплінах. Деякі нещодавні результати, де такі процеси використовуються для аналізу випадкових дерев, можна знайти в роботах [34, 36, 37, 41, 42, 50, 67]. Покажемо на прикладі як саме загальні гіллясті процеси пов'язані з випадковими деревами.

Нагадаємо, що випадкове рекурсивне дерево будується наступним чином. В нульовий момент часу присутній лише корінь. В кожний наступний дискретний момент часу навмання обирається одна із вершин дерева та до неї додається один син. Тепер розширимо цю модель на випадок неперервного часу. Аналогічно в нульовий момент часу присутній лише корінь. Нехай з коренем асоційовано таймер, що має експоненційний розподіл з інтенсивністю 1. Коли він закінчується, то до кореня додається нова вершина, таймер в корені перезапускається, а з новою вершиною асоціюється новий незалежний таймер з тим самим розподілом. Надалі, коли в певній вершині спрацьовує

таймер – до неї додається нова вершина аналогічним чином. Легко побачити, що час додавання другої вершини (коли присутній лише корінь) має експоненційний розподіл з інтенсивністю 1, а час між додаванням $(n + 1)$ -ої та n -ої вершини вже має експоненційний розподіл з інтенсивністю n , оскільки незалежно працюють n таймерів. З властивостей експоненційного розподілу також випливає, що вершина, де спрацює наступний таймер, розподілена рівномірно серед усіх присутніх вершин. Еволюція такого дерева повністю відповідає еволюції випадкового рекурсивного дерева.

В термінах процесів Крампа-Мода-Ягерса точковий процес матиме вигляд $Z = \sum_{n \geq 1} \delta_{S_n}$, де S_n – випадкове блукання зі стрибками ξ , що мають експоненційний розподіл з інтенсивністю 1. Саме при дослідженні цього процесу в роботі [36] встановлено функціональну збіжність профілю рекурсивного дерева на деякому фіксованому рівні, а в роботі [37] аналогічним чином вже досліджено профіль у випадку, коли рівень зростає разом з кількістю вершин.

Вищеописану модель легко узагальнити – для цього достатньо розглядати інші розподіли ξ , зберігши припущення, що вхідний точковий процес породжений випадковим блуканням. Саме для таких моделей, а також у випадку збурених випадкових блукань, проводяться дослідження в другому розділі дисертаційної роботи. Його мета складається з трьох частин. По-перше, досліджуються згортки функцій лінійного росту, а саме нас цікавлять згортки функцій відновлення. Саме такі згортки природно виникають при дослідженні наступних поколінь в гіллястих процесах, й вже впливали в вищеназваних роботах за простіших умов. По-друге, досліджена висота дерева загального гіллястого процесу. Й нарешті, всі результати перевіряються за допомогою математичного моделювання.

Огляд за розділом 3

В цьому розділі досліджуються асимптотичні властивості VP-дерев², проте їх огляд слід почати з практичної задачі, де вони виникають.

Мабуть одним із найбільш поширених типів дерев, що зустрічаються на практиці, є бінарні дерева пошуку, вони є і одними із найдослідженіших класів дерев. Вперше вони виникають як структура даних для швидкого доступу до елемента в масиві за ключем, де елементи зберігаються як пара ключ-значення. Необхідною умовою для їх побудови є лише можливість порівнювати ключі між собою, а основним нюансом пошуку є те, що шукається елемент, що в точності збігається з запитом. Проте нерідко виникають ситуації, коли запит просто не може бути точним, саме тоді виникає так звана задача пошуку найближчого сусіда (інколи в літературі “NN-пошук” та “нечіткий пошук”), коли замість точної рівності за ключем в множині шукається елемент з ключем, що найбільш близький чи схожий на заданий, а під поняттям схожості зазвичай розуміють відстань між ключами в деякому метричному просторі. Ця задача також нерідко розширюється до задачі пошуку $k \geq 2$ найближчих сусідів (kNN-пошук) або пошуку у радіусі, коли шукаються всі елементи, що лежать на відстані не більше ніж $r > 0$ від заданого.

На сьогодні пошук найближчого сусіда використовується в широкому колі задач, зокрема, як складова частина багатьох алгоритмів. Ось короткий список його застосувань:

- Географічний пошук використовується людьми кожен день, коли на карті шукаються місця чи заклади близькі до вашого GPS розташування.
- Пошук найближчих точок в обчислювальній геометрії. Наприклад, рутинною задачею в комп’ютерній графіці є пошук перетинів об’єктів, котрий можна оптимізувати, залишаючи для розгляду лише об’єкти, що розташовані близько.

²Скорочено від англ. vantage point tree.

- Аналогічно, пошук близьких об'єктів використовується для оптимізації підрахунку їх взаємодії в комп'ютерній симуляції фізичних процесів, коли взаємодією між дальніми об'єктами можна знехтувати.
- Пошук подібностей наразі присутній майже в усіх пошукових сервісах. Це може бути пошук схожих зображень або пошук подібних наукових статей та робіт до заданої, зокрема, перевірка на антиплагіат.
- kNN-пошук є важливим інструментом в машинному навчанні, зокрема:
 - Він використовується в різноманітних моделях регресії. Ми посилаємось на роботи [18] та [20] для детального представлення цих моделей.
 - Ми також посилаємось на [19], що є гарним викладом ідей, на яких базуються алгоритми класифікації, що використовують пошук найближчого сусіда.
 - В останніх дослідженнях почали комбінувати пошук найближчого сусіда та нейронні мережі для більш точних та більш стійких результатів, дивись, наприклад, [63].

Ці моделі виявились корисними в таких сферах класифікації як категоризація текстів, категоризація та пошук мультимедіа, розпізнавання образів, пошук інформації та у багато інших.

- Пошук найближчого сусіда є складовою різних евристик. Наприклад, з його допомогою можна швидко, але не точно, розв'язувати задачу комівояжера.

Зважаючи на актуальність цієї задачі, постає питання саме швидкого виконання запитів пошуку, що в основному досягається за допомогою використання спеціальних структур даних. Серед цих структур можна виділити KD-дерева [6, 59], дерева куль (або метричні дерева) [62] та VP-дерева. Остання структура була вперше представлена Янілосом у 1993 році в [79] та є об'єктом досліджень третього розділу дисертаційної роботи.

Хоча VP-дерева існують вже достатньо давно, наскільки нам відомо, вони не досліджувались з теоретичної точки зору, а наявні результати були отримані емпіричним шляхом. Серед присутніх у літературі результатів щодо VP-дерев можна виділити наступні:

- В роботах [31, 52, 56, 73] VP-дерева статистично досліджені на конкретних наборах даних, з чого автори роблять висновки щодо їх асимптотичної поведінки в загальному.
- Іншим підходом є порівняння різних структур дерев, наприклад, VP-дерев та KD-дерев, та зіставлення їх характеристик на однакових наборах даних, наприклад, дивись [49, 60]. З цього робляться висновки про ефективність однієї структури у порівнянні з іншою за певних умов.
- В роботах [51, 61, 73, 74] до моделі VP-дерева вносяться евристики та аналізується їх вклад та ефективність на специфічних даних.
- В роботах [13, 61, 77] розглядається поведінка та можливість використання VP-дерев у випадках, коли вхідні дані не задовольняють необхідним умовам, наприклад, не виконуються всі аксіоми метричного простору.

На відміну від усіх цих робіт ми використовуємо ймовірнісний підхід для аналізу VP-дерев та деяких його характеристик, коли кількість вершин у дереві зростає до нескінченності. Ми досліджуємо ймовірнісні властивості VP-дерев, а саме розглядаємо VP-дерево, що побудоване на послідовності незалежних однаково розподілених випадкових величин на $[-1, 1]^d$ з функцією відстані ℓ_∞ . Ми аналізуємо довжину найлівішого шляху в дереві та розбиття простору, що він створює. Основні результати включають декілька граничних теорем щодо збіжності цих характеристик, коли кількість вершин в дереві зростає до нескінченності. Наприкінці ми використовуємо отримані результати, щоб висунути гіпотезу про балансування випадкового VP-дерева, та виконуємо математичне моделювання як для її перевірки, так і для підтвердження інших результатів та демонстрації додаткових зображень еволюцій та закономірностей серед досліджуваних характеристик.

Розділ 1

Випадкові дерева простої будови

1.1 Означення та обговорення

1.1.1 Однорозгалуджене гіллясте випадкове блукання. Розглянемо систему частинок, що живуть на решітці \mathbb{Z} . В момент часу 0 в системі знаходиться одна частинка, що зазвичай розташована в 0. Система еволюціонує наступним шляхом: в кожний дискретний момент часу одна з частинок обирається рівномірно (якщо вона існує). Ця частинка (матір) замінюється на випадковий кластер частинок (дітей), чії відносні зміщення задаються точковим процесом $\zeta = \sum_{i=1}^N \delta_{Z_i}$. Тобто, якщо матір знаходилась в позиції x , тоді її діти знаходяться в позиціях $x + Z_1, \dots, x + Z_N$. Ми припускаємо, що число N є м.н. скінченним. Всі частинки діють незалежно. Послідовність точкових процесів на \mathbb{Z} , що описує місцезнаходження частинок впродовж поколінь, називається однорозгалудженим гіллястим випадковим блуканням (однорозгалудженим ГВБ).

Позначимо через S_n кількість частинок в момент часу n , та через $x_{1,n}, \dots, x_{S_n,n}$ – їх позиції. Профілем однорозгалудженого ГВБ в момент часу n є наступна функція:

$$\mathbb{L}_n(k) = \#\{1 \leq j \leq S_n : x_{j,n} = k\}, \quad n \in \mathbb{N}, \quad k \in \mathbb{Z}, \quad (1.1)$$

тобто $\mathbb{L}_n(k)$ – це кількість частинок в положенні $k \in \mathbb{Z}$ в момент часу n .

Позначимо через ν_k очікувану кількість частинок в положенні $k \in \mathbb{Z}$ в кластерному процесі ζ :

$$\nu_k = \mathbb{E}\zeta(\{k\}) = \mathbb{E} \left[\sum_{i=1}^N \mathbb{1}_{\{Z_i=k\}} \right], \quad k \in \mathbb{Z}.$$

Далі позначимо через $m(\beta)$ твірну функцію моментів інтенсивності кластерного процесу ζ мінус 1:

$$m(\beta) = \sum_{k \in \mathbb{Z}} e^{\beta k} \nu_k - 1 = \mathbb{E} \left[\sum_{i=1}^N e^{\beta Z_i} \right] - 1.$$

Припустимо, що існує відкритий інтервал \mathcal{I} , який містить 0, де функція m скінченна. Тоді m коректно визначена при $\beta \in \{z \in \mathbb{C} : \operatorname{Re} z \in \mathcal{I}\}$ та є строго опуклою й нескінченно диференційовною на \mathcal{I} . Визначимо функцію

$$\varphi(\beta) = \frac{m(\beta)}{m(0)}, \quad \operatorname{Re} \beta \in \mathcal{I}.$$

Також позначимо через $(\beta_-, \beta_+) \subset \mathcal{I}$ відкритий інтервал, в якому $\varphi'(\beta)\beta < \varphi(\beta)$:

$$\beta_- = \inf\{\beta \in \mathcal{I} : \varphi'(\beta)\beta < \varphi(\beta)\}, \quad (1.2)$$

$$\beta_+ = \sup\{\beta \in \mathcal{I} : \varphi'(\beta)\beta < \varphi(\beta)\}. \quad (1.3)$$

Інтервал (β_-, β_+) містить 0 й таким чином непорожній.

Нехай ν є мірою інтенсивності точкового процесу ζ :

$$\nu = \sum_{k \in \mathbb{Z}} \nu_k \delta_k.$$

(Нормалізована) твірна функція моментів профілю однорозгалудженого ГВБ визначена, при $\operatorname{Re} \beta \in \mathcal{I}$, формулою

$$W_n(\beta) = \frac{1}{n^{\varphi(\beta)}} \sum_{i=1}^{S_n} e^{\beta x_{i,n}}. \quad (1.4)$$

За певних умов (дивись розділ 1.2) W_n збігається при $n \rightarrow \infty$ до випадкової аналітичної функції W_∞ з ймовірністю 1. Зауважимо, що $W_\infty(0) = m(0)$, оскільки $W_n(0) = S_n/n$.

Нарешті, визначимо наступні випадкові величини:

$$\chi_1(0) = (\log W_\infty)'(0) = \frac{W'_\infty(0)}{m(0)}, \quad (1.5)$$

$$\chi_2(0) = (\log W_\infty)''(0) = \frac{W''_\infty(0)m(0) - (W'_\infty(0))^2}{m^2(0)} \quad (1.6)$$

та, більш загально, $\chi_j(\beta) = (\log W_\infty)^{(j)}(\beta)$ ($j \geq 1$).

1.1.2 Моделі випадкових дерев. Випадкове дерево простої будови – дерево, що побудоване за наступним послідовним процесом. Всі вершини можуть бути двох типів: зовнішні та внутрішні. В момент часу 0 дерево зазвичай складається з лише однієї зовнішньої вершини, що є коренем дерева. В кожен дискретний момент часу рівноймовірно обирається одна зовнішня вершина, вона зазвичай стає внутрішньою та нові скінченні (можливо випадкові) піддерева приєднуються до цієї вибраної вершини та, можливо, сусідніх з нею.

Наведемо природну процедуру кодування еволюції випадкового дерева простої будови, що використовує однорозгалуджене ГВБ. В цьому кодуванні зовнішнім вершинам відповідають живі частинки, внутрішні вершини, що були обрані на деякому кроці, відповідають мертвим частинкам, та глибина, тобто відстань до кореня, кожної вершини – це позиція відповідної частинки.

Розглянемо основні класичні моделі випадкових дерев.

***D*-арні дерева ($D \geq 2$).**

В момент часу 0 *D*-арне дерево складається лише з його кореня. В кожен дискретний момент часу обирається одна випадкова зовнішня вершина й замінюється на внутрішню вершину з приєднаними *D* дітьми (зовнішніми вершинами). Іншими словами, всі зовнішні вершини *D*-арного дерева є його листками, та на кожному кроці додаються *D* синів до випадкового листка, обраного рівноймовірно.

У випадку, коли $D = 2$, дерево стає добре відомим бінарним деревом пошуку (БДП), котре також може бути побудоване з використанням послідовності рівномірно розподілених чисел.



Рис. 1.1: Правило побудови D -арного дерева при $D = 3$ та приклад його реалізації на момент часу $n = 3$.

Еволюція D -арного дерева може бути представлена як однорозгалуджене ГВБ з детермінованим точковим процесом для зміщень $\zeta = D\delta_1$. Маємо

$$\varphi(\beta) = \frac{De^\beta - 1}{D - 1}, \quad m(0) = D - 1, \quad \varphi^{(j)}(0) = \frac{D}{D - 1}, \quad j \in \mathbb{N}.$$

Константи β_- та β_+ є розв'язками рівняння $De^\beta(1 - \beta) = 1$.

Випадкові рекурсивні дерева.

Випадкове рекурсивне дерево (ВРД) будується за наступною процедурою. В момент часу 0 ВРД складається з однієї вершини – кореня. В кожен дискретний момент часу випадкова вершина, що існує, обирається та нова вершина приєднується до неї. В цьому представленні кожна вершина розглядається як зовнішня.

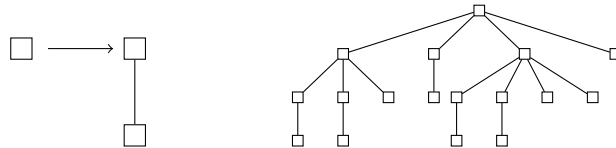


Рис. 1.2: Правило створення ВРД та приклад його реалізації на момент часу $n = 16$.

Однорозгалуджене ГВБ, що відповідає ВРД, будується з детермінованого точкового процесу для зміщень $\zeta = \delta_0 + \delta_1$, для котрого

$$\varphi(\beta) = e^\beta, \quad m(0) = 1, \quad \varphi^{(j)}(0) = 1, \quad j \in \mathbb{N}, \quad \beta_- = -\infty, \quad \beta_+ = 1.$$

p -орієнтовані рекурсивні дерева ($p \geq 2$).

p -орієнтоване рекурсивне дерево починається з однієї внутрішньої вершини (кореня) та однієї зовнішньої вершини, що до нього приєднана. Правило побу-

дови зображено на рисунку 1.3 (зліва) при $p = 2$. Для довільного $p = 2, 3, \dots$ правило аналогічне, проте з p замість двох зовнішніх вершин всередині.

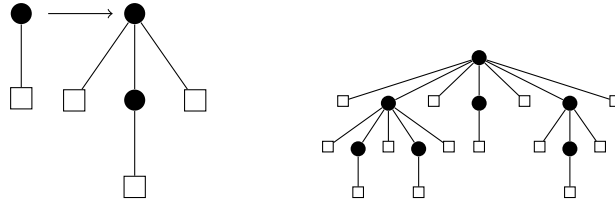


Рис. 1.3: Правило побудови p -орієнтованого рекурсивного дерева з $p = 2$ та приклад його реалізації на момент часу $n = 6$.

При $p = 2$ ця конструкція визначає добре відому конструкцію планарно-орієнтованого рекурсивного дерева (ПОРД).

Однорозгалуджене ГВБ, що відповідає p -орієнтованому рекурсивному дереву, починається з найпершої частки, що розташована в $x_{1,0} = 1$, та має детермінований точковий процес для зміщень $\zeta = p\delta_0 + \delta_1$, для котрого

$$\varphi(\beta) = \frac{1}{p}(e^\beta + p - 1), \quad m(0) = p, \quad \varphi^{(j)}(0) = \frac{1}{p}, j \in \mathbb{N}, \quad \beta_- = -\infty$$

та $\beta_+ \in$ розв'язком $e^\beta(\beta - 1) = p - 1$.

1.1.3 Функціонали на випадкових деревах. Існує багато різноманітних функціоналів, що означені на деревах. Ми зосереджуємо нашу увагу лише на деяких із них. Глибина (або рівень) вершини, за означенням, – це її відстань від кореня. Позначимо через $\mathbb{L}_n(k)$ кількість зовнішніх вершин дерева в момент часу n на глибині k . Це означення еквівалентне (за виключенням p -орієнтованих дерев) (1.1), визначеного для однорозгалудженого ГВБ з параметрами, що відповідають вибраному типу дерева. Цей функціонал називають зовнішнім профілем випадкового дерева (або число заповнення) на рівні k .

Мода u_n випадкового дерева визначається рівністю

$$u_n = \arg \max_{k \in \mathbb{N}_0} \mathbb{L}_n(k), \quad n \in \mathbb{N},$$

тобто u_n – це рівень, що містить найбільшу кількість зовнішніх вершин.

Ширину M_n випадкового дерева визначають наступною формулою

$$M_n = \max_{k \in \mathbb{N}_0} \mathbb{L}_n(k),$$

та вона рівна найбільшій кількості зовнішніх вершин на одному рівні.

1.2 Результати для випадкових дерев

Почнемо з леми, в якій перелічено деякі властивості однорозгалуджених гіллястих випадкових блукань, що відповідають вищеназваним трьом типам випадкових дерев.

Лема 1. Для однорозгалуджених гіллястих випадкових блукань, що відповідають D -арному дереву, випадковому рекурсивному дереву чи p -орієнтованому рекурсивному дереву, наступні твердження виконуються:

(а) $\nu_k > 0$ для хоча б одного $k \in \mathbb{Z} \setminus \{0\}$;

(б) кластерний точковий процес ζ м.н. непорожній та має хоча б дві частки з додатною ймовірністю;

(в) функція $t(\beta)$ скінченна на деякому непорожньому відкритому інтервалі \mathcal{I} , що містить 0;

(г) ν не сконцентровано на будь-якій власній адитивній підгрупі \mathbb{Z} ;

(д) для будь-якого $\beta \in (\beta_-, \beta_+)$ існує $\gamma = \gamma(\beta) > 1$ таке, що

$$\mathbb{E} \left[\left(\sum_{i=1}^N e^{\beta Z_i} \right)^\gamma \right] < \infty.$$

Доведення. (а) Для кожного процесу існує хоча б одна частинка δ_1 .

(б) Для всіх процесів існує хоча б дві частинки (рівно 2 для ВРД, D для D -арного дерева та $p + 1$ для p -орієнтованого рекурсивного дерева).

(в) Функція $t(\beta)$ є скінченною сумою експонент, вона скінченна на будь-якому скінченному інтервалі \mathcal{I} .

(г) Для кожного процесу існує хоча б одна частинка δ_1 , тому найбільший спільний дільник для величин зміщень рівний 1.

(д) Для кожного процесу сума під математичним сподіванням м.н. обмежена

невипадковою скінченною константою; γ може бути обране, наприклад, як $\gamma = 2$. \square

Ми навели цю лему для однорозгалуджених гіллястих випадкових блукань, що були представлені вище, проте вона справджується для набагато більшого класу однорозгалуджених ГВБ. Один із результатів, що випливає з леми 1, – це існування $W_\infty(\beta)$ такого, що $W_n(\beta) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} W_\infty(\beta)$ локально рівномірно на відкритому околі інтервалу (β_-, β_+) в \mathbb{C} , дивись [16, 43]. Відтепер ми припускаємо, що аргумент β завжди лежить в цій відкритій множині, де $W_\infty(\beta)$ коректно визначена.

Наступний результат дає характеристику випадкових величин $W_\infty(\beta)$ та $\chi_j(\beta)$ через стохастичні рівняння нерухомої точки.

Теорема 2. *Для однорозгалудженого ГВБ такого, що твердження леми 1 виконуються та точковий процес для зміщень ζ детермінований й складається з N точок, випадкова величина $W_\infty(\beta)$ задовольняє наступне стохастичне рівняння нерухомої точки*

$$W_\infty(\beta) \stackrel{d}{=} \sum_{i=1}^N e^{\beta Z_i} U_i^{\varphi(\beta)} W_\infty^{(i)}(\beta), \quad (1.7)$$

де $W_\infty^{(i)}(\beta)$ незалежні в сукупності копії $W_\infty(\beta)$, що не залежать від випадкового вектора $(U_i)_{i=1}^N$, що має розподіл Діріхле з параметром $\frac{1}{N-1}$.

Наступне доведення покладається на поняття урнної моделі Пойа-Еггенбергера. В цій моделі наявна єдина урна, що містить кульки N різних кольорів. Урна розвивається у дискретному часі наступним чином. На кожному кроці ми навмання обираємо кульку з урни, запам'ятовуємо колір цієї кульки $i \in \overline{1, N}$ та потім повертаємо її до урни. В залежності від кольору i витягнутої кульки ми також додаємо $a_{i,j}$ кульок кольору j до урни, для всіх $j \in \overline{1, N}$. Схема заміщення зазвичай представлена у вигляді матриці $A = (a_{i,j})_{i,j=1}^N$. Значення $a_{i,j}$ можуть бути нулями або навіть від'ємними. Наприклад, якщо $a_{2,4} = -3$, тоді ми забираємо 3 кульки кольору 4 з урни, коли витягаємо кульку кольору 2. Урна Пойа-Еггенбергера називається *витривалою* якщо можли-

во продовжувати процес еволюції урни нескінченно на кожному можливому стохастичному шляху. Якщо всі $a_{i,j}$ невід’ємні, тоді відповідна модель урни очевидно витривала. Проте, якщо деяке $a_{i,j}$ від’ємне, то можлива ситуація, коли за правилом заміщення ми матимемо забрати $|a_{i,j}|$ кульок кольору j , котрих просто не вистачає в урні. Витривалі урни та властивості витривалості є найбільш дослідженими для моделі урни Пойа-Еггенбергера. Для більшої інформації відносно моделей урн та деяких класичних результатів ми посилаємось на книгу [55].

Доведення теореми 2. Факт, що точковий процес ζ детермінований означає, що кількість частинок N , а також їх позиції Z_i є константами. З леми 1 легко випливає, що $N \geq 2$.

В момент часу $n = 0$ присутня лише одна частинка, розташована в $x_{1,0}$. В момент часу $n = 1$ вона розбивається на N частинок, що розташовані в $x_{1,0} + Z_1, \dots, x_{1,0} + Z_N$. Кожна частинка тепер еволюціонує незалежно та її еволюція така ж сама, як і у початкової частинки.

Уявимо однорозгалуджене ГВБ в момент часу $n + 1$ ($n > 0$). Можна переписати (1.4), використовуючи значення $W_k^{(i)}(\beta)$ ($k \in \overline{1, N}$) підпроцесів на частинках, що були створені після найпершого розбиття, наступним чином

$$W_{n+1}(\beta) \stackrel{d}{=} \frac{1}{(n+1)^{\varphi(\beta)}} \sum_{i=1}^N \left(e^{\beta Z_i} (V_n^{(i)})^{\varphi(\beta)} W_{V_n^{(i)}}^{(i)}(\beta) \mathbb{1}_{\{V_n^{(i)} > 0\}} + e^{\beta(x_{1,0} + Z_i)} \mathbb{1}_{\{V_n^{(i)} = 0\}} \right),$$

де $V_n^{(i)}$ – це число подій розбиття, що відбулись у підпроцесі, що був породжений частинкою $x_{1,0} + Z_i$ (що була породжена після першого розбиття) та, оскільки n подій розбиття відбулось в загальному:

$$\sum_{i=1}^N V_n^{(i)} = n.$$

За симетричністю всі $V_n^{(i)}$ мають однаковий розподіл для $i = 1, \dots, N$, оскільки підпроцеси є незалежними та однаково розподіленими. Також $V_n^{(i)} \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \infty$ та, отже, $W_{V_n^{(i)}}^{(i)}(\beta) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} W_\infty^{(i)}(\beta)$ для $i = 1, \dots, N$.

Залишається лише знайти граничний розподіл $V_n^{(i)}/(n+1)$ при n , що прямує до нескінченності. Він може бути переписаний наступним шляхом:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{V_n^{(i)}}{n+1} &= \lim_{n \rightarrow \infty} \frac{V_n^{(i)}}{n} = \lim_{n \rightarrow \infty} \frac{V_n^{(i)}(N-1)}{n(N-1)} = \lim_{n \rightarrow \infty} \frac{1+V_n^{(i)}(N-1)}{N+n(N-1)} = \\ &= \lim_{n \rightarrow \infty} \frac{S_n^{(i)}}{S_n^*}, \end{aligned}$$

де $S_n^{(i)}$ позначає суму частинок в i -му підпроцесі та $S_n^* = S_{n+1}$ – сума частинок в усіх підпроцесах.

Зауважимо, що кожна подія розбиття відбувається з того, що обирається випадкова частинка з усіх присутніх частинок, тоді $N-1$ частинок додаються до підпроцесу, що відповідає обраній частинці. Це означає, що ймовірність, що подія розбиття відбудеться в i -му підпроцесі пропорційна поточній кількості частинок в ньому. Наведений процес може бути описаний в термінах моделі урни Пойа-Еггенбергера, де на початку були присутні рівно по одній кульці кожного кольору та схема заміщення має вигляд $A = (a_{i,j})_{i,j=1}^N$ з $a_{i,i} = N-1$ та $a_{i,j} = 0$ при $i \neq j$. Тут кульки кольору i відповідають частинкам в i -му підпроцесі. Це витривала модель урни, для котрої Гоует довів наступний результат, дивись [33],

$$\lim_{n \rightarrow \infty} \frac{S_n^{(i)}}{S_n^*} = U_i,$$

де вектор $(U_i)_{i=1}^N$ має розподіл Діріхле з параметром $\frac{1}{N-1}$.

Рівняння (1.7) випливає з поєднання вищенаведених формул. □

Зауваження 3. Випадковий вектор (x_1, x_2, \dots, x_N) має розподіл Діріхле з параметром $\alpha > 0$, якщо $x_i \geq 0$, $i = 1, \dots, N$, $\sum_{i=1}^N x_i = 1$ та функція щільності відносно $(N-1)$ -вимірної міри Лебега задана наступним чином

$$f(x_1, \dots, x_N, \alpha) = \frac{\Gamma(\alpha N)}{\Gamma(\alpha)^N} \prod_{i=1}^N x_i^{\alpha-1},$$

де $\Gamma(\alpha)$ позначає гамма-функцію.

Наслідок 4. За припущень теореми 2 випадкова величина $\chi_1(0)$ задовольняє стохастичне рівняння нерухомої точки

$$\chi_1(0) \stackrel{d}{=} \sum_{i=1}^N (Z_i U_i + \varphi'(0) U_i \log U_i + U_i \chi_1^{(i)}(0)), \quad (1.8)$$

де $\chi_1^{(i)}(0)$ – незалежні в сукупності копії $\chi_1(0)$, що не залежать від випадкового вектора $(U_i)_{i=1}^N$, який має розподіл Діріхле з параметром $\frac{1}{N-1}$.

Доведення. Нагадаємо, що $W_\infty(0) = m(0)$ та $\varphi(0) = 1$. Взявши логарифмічну похідну в (1.7) та підставивши $\beta = 0$, отримуємо

$$\frac{W'_\infty(0)}{m(0)} \stackrel{d}{=} \frac{\sum_{i=1}^N (Z_i U_i m(0) + \varphi'(0) U_i \log U_i m(0) + U_i W'_\infty(0))}{\sum_{i=1}^N U_i m(0)}.$$

З означення (1.5) величини $\chi_1(0)$ випливає, що $W'_\infty(0) = m(0)\chi_1(0)$. За означенням розподілу Діріхле маємо $\sum_{i=1}^N U_i = 1$. Поєднавши ці рівності, отримуємо (1.8). \square

Зауваження 5. Використавши ті ж міркування, що й у доведенні наслідку 4, можна також вивести аналогічні рівняння для $\chi_j(\beta)$.

Приклад 6. Підрахуємо наведені величини для БДП, тобто для D -арного дерева з $D = 2$, дивись підрозділ 1.1.2. З теореми 2 ми отримуємо стохастичне рівняння нерухомої точки для БДП:

$$W_\infty(\beta) \stackrel{d}{=} e^\beta U_1^{2e^\beta - 1} W_\infty^{(1)}(\beta) + e^\beta U_2^{2e^\beta - 1} W_\infty^{(2)}(\beta). \quad (1.9)$$

Вектор (U_1, U_2) має розподіл Діріхле з параметром 1. Він може бути зображений як $(\frac{Y_1}{Y_1+Y_2}, \frac{Y_2}{Y_1+Y_2})$, де Y_1, Y_2 незалежні випадкові величини зі стандартним експоненційним розподілом, дивись [21]. Тоді для $x \in [0, 1]$:

$$\begin{aligned} \mathbb{P} \left\{ \frac{Y_1}{Y_1 + Y_2} \leq x \right\} &= \mathbb{P} \left\{ Y_1 \leq \frac{x Y_2}{1 - x} \right\} = \int_0^\infty e^{-y_2} \int_0^{\frac{x y_2}{1-x}} e^{-y_1} dy_1 dy_2 = \\ &= \int_0^\infty e^{-y_2} \left(1 - e^{-\frac{x y_2}{1-x}} \right) dy_2 = x. \end{aligned}$$

Це означає, що випадкова величина $\frac{Y_1}{Y_1+Y_2}$ має стандартний рівномірний розподіл на $[0, 1]$, отже, вектор $(\frac{Y_1}{Y_1+Y_2}, \frac{Y_2}{Y_1+Y_2})$ має той самий розподіл, що й вектор $(U, 1 - U)$, де U має рівномірний розподіл на $[0, 1]$.

Отже, ми можемо переписати (1.9) таким чином

$$W_\infty(\beta) \stackrel{d}{=} e^\beta U^{2e^\beta - 1} W_\infty^{(1)}(\beta) + e^\beta (1 - U)^{2e^\beta - 1} W_\infty^{(2)}(\beta).$$

Згідно з наслідком 4 ми також маємо стохастичне рівняння нерухомої точки

$$\chi_1(0) \stackrel{d}{=} 1 + 2U \log U + U \chi_1^{(1)}(0) + 2(1 - U) \log(1 - U) + (1 - U) \chi_1^{(2)}(0).$$

Це добре відоме рівняння нерухомої точки, яке визначає (з точністю до адитивної константи) граничний розподіл для числа порівнянь в алгоритмі швидкого сортування, дивись [66].

1.2.1 Результати для зовнішніх профілів випадкових дерев.

Теорема 7. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D - 1)n + 1$ зовнішніми вершинами. Для всіх $r \in \mathbb{N}_0$ ми маємо

$$(\log n)^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \left| \frac{1}{n} \mathbb{L}_n(k) - H_n(k) \sum_{j=0}^r G_j \left(\frac{k - \frac{D}{D-1} \log n}{\sqrt{\frac{D}{D-1} \log n}}; 0 \right) \frac{1}{(\log n)^{j/2}} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де G_j наведено в зауваженні 10 нижче, та

$$H_n(k) = \frac{(D - 1)^{3/2}}{\sqrt{2D\pi \log n}} e^{-\frac{(k - \frac{D}{D-1} \log n)^2}{\frac{2D}{D-1} \log n}}.$$

Ця теорема може бути легко порівняна з наступною центральною граничною теоремою, що була доведена Шавен, Дрмотою та Джабур-Хаттабом для БДП, дивись [15],

$$\sup_{k \in \mathbb{Z}} \left| \frac{1}{n} \mathbb{L}_n(k) - \frac{1}{\sqrt{4\pi \log n}} e^{-\frac{(k - 2 \log n)^2}{4 \log n}} \right| = O\left(\frac{1}{\log n}\right) \quad \text{м.н.}$$

Цей результат випливає з теореми 7 при $D = 2$ та $r = 0$, проте вищенаведене твердження дає набагато більш повний результат.

Теорема 8. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – профіль випадкового рекурсивного дерева з $n + 1$ вершин. Для всіх $r \in \mathbb{N}_0$ ми маємо

$$(\log n)^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \left| \frac{1}{n} \mathbb{L}_n(k) - \frac{e^{-\frac{(k - \log n)^2}{2 \log n}}}{\sqrt{2\pi \log n}} \sum_{j=0}^r G_j \left(\frac{k - \log n}{\sqrt{\log n}}; 0 \right) \frac{1}{(\log n)^{j/2}} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де G_j наведено в зауваженні 10 нижче.

Теорема 9. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль p -орієнтованого рекурсивного дерева з $pn + 1$ зовнішньою вершиною. Для всіх $r \in \mathbb{N}_0$ маємо

$$(\log n)^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \left| \frac{1}{n} \mathbb{L}_n(k) - H_n(k) \sum_{j=0}^r G_j \left(\frac{k - 1 - \frac{1}{p} \log n}{\sqrt{\log n/p}}; 0 \right) \frac{1}{(\log n)^{j/2}} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де G_j наведено в зауваженні 10 нижче та

$$H_n(k) = \frac{p^{3/2}}{\sqrt{2\pi \log n}} e^{-\frac{p(k-1-\frac{1}{p} \log n)^2}{2 \log n}}.$$

Доведення теорем 7, 8, 9. Ці результати отримуються з теореми 59, наведеної в додатку для однорозгалуженого ГВБ, що відповідає випадковим деревам як описано в розділі 1.1.2, та леми 1. Для цього в теоремі необхідно покласти $\beta = 0$. \square

Зауваження 10. $G_j(x; \beta)$ – це поліном степеня не вище ніж $3j$, що заданий як

$$G_j(x; \beta) = \frac{(-1)^j}{j!} e^{\frac{1}{2}x^2} B_j(D_1, \dots, D_j) e^{-\frac{1}{2}x^2},$$

де B_j – j -й (експоненційний) повний поліном Белла, дивись [17, глава 3.3], та D_1, D_2, \dots позначають диференціальні оператори з випадковими коефіцієнтами, що задані формулою

$$D_j = \frac{\varphi^{(j+2)}(\beta)}{(j+1)(j+2)} \left(\frac{1}{\sqrt{\varphi''(\beta)}} \frac{d}{dx} \right)^{j+2} + \chi_j(\beta) \left(\frac{1}{\sqrt{\varphi''(\beta)}} \frac{d}{dx} \right)^j.$$

Зокрема,

$$G_0(x; 0) = 1$$

та $G_j(0; \beta) = 0$ для непарних j .

Теорема 11. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D-1)n + 1$ зовнішніми вершинами. Для всіх $r \in \mathbb{N}_0$ та кожної компакної множини $L \subset (\frac{D}{D-1}e^{\beta-}, \frac{D}{D-1}e^{\beta+})$ виконується

$$(\log n)^{r+1} \sup_{k \in \mathbb{Z} \cap (\log n)L} \left| H_n(k) \mathbb{L}_n(k) - \frac{1}{\sqrt{2\pi k}} \sum_{j=0}^r \frac{F_{2j}(0; \log \frac{(D-1)k}{D \log n})}{(\log n)^j} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де F_{2j} наведено в зауваженні 14 та

$$H_n(k) = n^{\frac{1}{D-1}} \left(\frac{(D-1)k}{De \log n} \right)^k. \quad (1.10)$$

Теорема 12. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – профіль випадкового рекурсивного дерева з $n+1$ вершиною. Для всіх $r \in \mathbb{N}_0$ та кожної компактної множини $L \subset (0, e)$ виконується

$$(\log n)^{r+1} \sup_{k \in \mathbb{Z} \cap (\log n)L} \left| \left(\frac{k}{e \log n} \right)^k \mathbb{L}_n(k) - \frac{1}{\sqrt{2\pi k}} \sum_{j=0}^r \frac{F_{2j}(0; \log \frac{k}{\log n})}{(\log n)^j} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де F_{2j} наведено в зауваженні 14.

Теорема 13. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль p -орієнтованого рекурсивного дерева з $pn+1$ зовнішньою вершиною. Для всіх $r \in \mathbb{N}_0$ та кожної компактної множини $L \subset (0, \frac{1}{p}e^{\beta+})$ виконується

$$(\log n)^{r+1} \sup_{k-1 \in \mathbb{Z} \cap (\log n)L} \left| \mathbb{L}_n(k) - \sum_{j=0}^r \frac{F_{2j}(0; \log \frac{p(k-1)}{\log n})}{\sqrt{2\pi(k-1)} (\log n)^j} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де F_{2j} наведено в зауваженні 14 та

$$H_n(k) = n^{\frac{1-p}{p}} \left(\frac{p(k-1)}{e \log n} \right)^{k-1}. \quad (1.11)$$

Доведення теорем 11, 12, 13. Ці результати випливають з леми 1 та теорем 59, наведеній у додатку, для однорозгалуджених ГВБ, що відповідають випадковим деревам, як було описано в розділі 1.1.2.

Ідея результату полягає в тому, щоб обирати залежність між k та β таким чином, щоб значення x в поліномах $G(x; \beta)$ дорівнювало нулеві, тобто щоб виконувалось співвідношення

$$k - \varphi'(\beta) \log n = 0.$$

Це рівняння розв'язується відносно невідомого β , що, наприклад, для D -арних дерев рівне $\log \frac{(D-1)k}{D \log n}$. З іншого боку цей розв'язок все ще має належати області збіжності, тобто виникає умова

$$k \in (\log n \varphi'(\beta_-), \log n \varphi'(\beta_+)).$$

Зауважимо, що за такого вибору параметрів, $G_j(x; \beta)$ обертається в нуль при непарних значеннях j . \square

Зауваження 14. Якщо покласти $F_{2j}(0; \beta) = W_\infty(\beta)G_{2j}(0; \beta)$, то можна побачити, що $F_{2j}(0; \beta)$ є лінійною комбінацією $1, W_\infty(\beta), \dots, W_\infty^{(2j)}(\beta)$. Зокрема

$$F_0(0; \beta) = W_\infty(\beta).$$

У наступних теоремах наведена поведінка профілів у випадках, коли рівень, що розглядається, також зростає з n .

Теорема 15. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль D -арного дерева з $(D-1)n+1$ зовнішньою вершиною. Покладемо $\mathbb{L}_n^\circ(k) := \mathbb{L}_n(k) - \mathbb{E}[\mathbb{L}_n(k)]$ та нехай $(k_n)_{n \in \mathbb{N}}$ позначає детерміновану цілочисельну послідовність.

(а) Якщо $k_n = \frac{D}{D-1} \log n + \alpha \sqrt{\frac{D}{D-1} \log n} + o(\sqrt{\log n})$ для деякого $\alpha \in \mathbb{R}$, тоді

$$\frac{\log n}{n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)^2 \alpha e^{-\frac{1}{2}\alpha^2}}{D\sqrt{2\pi}} (\chi_1(0) - \mathbb{E}[\chi_1(0)]).$$

(б) Якщо $k_n = \frac{D}{D-1} \log n + c_n$, де $c_n = o(\log n)$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n e^{c_n}} \left(\frac{(D-1)k_n}{D \log n} \right)^{k_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)(\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\left(\frac{D}{D-1}\right)^{3/2} \sqrt{2\pi}}.$$

Зокрема, якщо $c_n = o(\sqrt{\log n})$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{(D-1)(\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\left(\frac{D}{D-1}\right)^{3/2} \sqrt{2\pi}}.$$

(в) Якщо $k_n = \frac{D}{D-1} \log n + c_n$, де $c_n = O(1)$, тоді

$$\frac{(\log n)^{3/2}}{n} \mathbb{L}_n^\circ(k_n) - \frac{D-1}{\left(\frac{D}{D-1}\right)^{3/2} \sqrt{2\pi}} (R(c_n) - \mathbb{E}[R(c_n)]) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

$$\text{де } R(c_n) = (c_n + \frac{1}{2})\chi_1(0) - \frac{\chi_1^2(0) - \chi_2(0)}{2}.$$

Теорема 16. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – профіль випадкового рекурсивного дерева з $n + 1$ вершиною. Покладемо $\mathbb{L}_n^\circ(k) := \mathbb{L}_n(k) - \mathbb{E}[\mathbb{L}_n(k)]$ та нехай $(k_n)_{n \in \mathbb{N}}$ позначає детерміновану цілочисельну послідовність.

(а) Якщо $k_n = \log n + \alpha\sqrt{\log n} + o(\sqrt{\log n})$ для деякого $\alpha \in \mathbb{R}$, тоді

$$\frac{\log n}{n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{\alpha e^{-\frac{1}{2}\alpha^2}}{\sqrt{2\pi}} (\chi_1(0) - \mathbb{E}[\chi_1(0)]).$$

(б) Якщо $k_n = \log n + c_n$, де $c_n = o(\log n)$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n e^{c_n}} \left(\frac{k_n}{\log n} \right)^{k_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{\chi_1(0) - \mathbb{E}[\chi_1(0)]}{\sqrt{2\pi}}.$$

Зокрема, якщо $c_n = o(\sqrt{\log n})$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{\chi_1(0) - \mathbb{E}[\chi_1(0)]}{\sqrt{2\pi}}.$$

(в) Якщо $k_n = \log n + c_n$, де $c_n = O(1)$, тоді

$$\frac{(\log n)^{3/2}}{n} \mathbb{L}_n^\circ(k_n) - \frac{1}{\sqrt{2\pi}} (R(c_n) - \mathbb{E}[R(c_n)]) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де $R(c_n) = (c_n + \frac{1}{2})\chi_1(0) - \frac{\chi_1^2(0) - \chi_2(0)}{2}$.

Теорема 17. Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) – зовнішній профіль p -орієнтованого рекурсивного дерева з $pn + 1$ зовнішньою вершиною. Покладемо $\mathbb{L}_n^\circ(k) := \mathbb{L}_n(k) - \mathbb{E}[\mathbb{L}_n(k)]$ та нехай $(k_n)_{n \in \mathbb{N}}$ позначає детерміновану цілочисельну послідовність.

(а) Якщо $k_n = \frac{1}{p} \log n + 1 + \alpha\sqrt{\frac{\log n}{p}} + o(\sqrt{\log n})$ для деякого $\alpha \in \mathbb{R}$, тоді

$$\frac{\log n}{n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{p^2 \alpha e^{-\frac{1}{2}\alpha^2}}{\sqrt{2\pi}} (\chi_1(0) - \mathbb{E}[\chi_1(0)]).$$

(б) Якщо $k_n = \frac{1}{p} \log n + 1 + c_n$, де $c_n = o(\log n)$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n e^{c_n}} \left(\frac{pk_n}{\log n} \right)^{k_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{p^{5/2} (\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\sqrt{2\pi}}.$$

Зокрема, якщо $c_n = o(\sqrt{\log n})$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{nc_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{p^{5/2}(\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\sqrt{2\pi}}.$$

(в) Якщо $k_n = \frac{1}{p} \log n + 1 + c_n$, де $c_n = O(1)$, тоді

$$\frac{(\log n)^{3/2}}{n} \mathbb{L}_n^\circ(k_n) - \frac{p^{5/2}}{\sqrt{2\pi}} (R(c_n) - \mathbb{E}[R(c_n)]) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

$$\text{де } R(c_n) = (c_n + \frac{1}{2})\chi_1(0) - \frac{\chi_1^2(0) - \chi_2(0)}{2}.$$

Доведення теорем 15, 16, 17. Ці результати є наслідками теореми 60 з додатку для однорозгалуджених ГВБ, що відповідають випадковим деревам як описано в розділі 1.1.2, та леми 1. \square

Зауваження 18. Зауважимо, що всі результати для p -орієнтованих рекурсивних дерев зміщені на $x_{1,0} = 1$, що є позицією початкової частинки у відповідному однорозгалудженому ГВБ.

1.2.2 Результати щодо моди випадкових дерев.

Теорема 19. *Існує м.н. скінченна випадкова величина K така, що для всіх $n > K$ мода u_n для D -арного дерева з $(D - 1)n + 1$ зовнішньою вершиною рівна одному з чисел*

$$\left\lfloor \frac{D}{D-1} \log n + \chi_1(0) - \frac{1}{2} \right\rfloor \text{ або } \left\lceil \frac{D}{D-1} \log n + \chi_1(0) - \frac{1}{2} \right\rceil.$$

Теорема 20. *Існує м.н. скінченна випадкова величина K така, що для всіх $n > K$ мода u_n для випадкового рекурсивного дерева $n + 1$ вершиною рівна одному з чисел*

$$\left\lfloor \log n + \chi_1(0) - \frac{1}{2} \right\rfloor \text{ або } \left\lceil \log n + \chi_1(0) - \frac{1}{2} \right\rceil.$$

Теорема 21. *Існує м.н. скінченна випадкова величина K така, що для всіх $n > K$ мода u_n для p -орієнтованого рекурсивного дерева з $pn + 1$ зовнішньою вершиною рівна одному з чисел*

$$\left\lfloor \frac{1}{p} \log n + \chi_1(0) + \frac{1}{2} \right\rfloor \text{ або } \left\lceil \frac{1}{p} \log n + \chi_1(0) + \frac{1}{2} \right\rceil.$$

Доведення теорем 19, 20, 21. Ці результати є наслідками з теореми 61 в додатку для однорозгалуджених ГВБ, що відповідають випадковим деревам як описано в розділі 1.1.2, та леми 1. □

Зауваження 22. З доведення теореми 61 можна також зробити додаткові висновки щодо поведінки моди, а саме

- майже напевно мода не рівна одному з двох наведених в теоремах значень або не унікальна лише для скінченної кількості значень n ;
- майже напевне існують інтервали довільної довжини з послідовних n , де мода рівна саме першому значенню наведеному в теоремі (тобто округленню вниз); аналогічні інтервали існують і для другого значення.

1.2.3 Результати щодо ширини випадкових дерев.

Теорема 23. *Нехай M_n – ширина D -арного дерева з $(D - 1)n + 1$ зовнішньою вершиною, тоді виконується*

$$\frac{\sqrt{2D\pi \log n}}{(D - 1)^{3/2}n} M_n \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1.$$

Теорема 24. *Нехай M_n – ширина випадкового рекурсивного дерева з $n + 1$ вершиною, тоді виконується*

$$\frac{\sqrt{2\pi \log n}}{n} M_n \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1.$$

Вищенаведений результат є добре відомим в літературі, дивись [25]. Проте він був доведений зовсім іншими методами.

Теорема 25. *Нехай M_n – ширина p -орієнтованого рекурсивного дерева з $pn + 1$ зовнішньою вершиною, тоді виконується*

$$\frac{\sqrt{2\pi \log n}}{p^{3/2}n} M_n \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1.$$

Доведення теорем 23, 24, 25. Ці результати є наслідками теореми 62 в додатку для однорозгалуджених ГВБ, що відповідають випадковим деревам як описано в розділі 1.1.2, та леми 1. □

Теорема 26. Нехай M_n – ширина D -арного дерева з $(D - 1)n + 1$ зовнішньою вершиною. З ймовірністю 1 множина границь підпоследовностей послідовності

$$\tilde{M}_n := 2 \frac{D}{D-1} \log n \left(1 - \frac{\sqrt{2\pi \log n} \sqrt{\frac{D}{D-1}} M_n}{(D-1)n} \right), \quad n \in \mathbb{N},$$

є інтервалом $[\chi_2(0) - 1/12, \chi_2(0) + 1/6]$. Більш того, якщо покласти $\theta_n := \min_{k \in \mathbb{Z}} \left| \frac{D}{D-1} \log n + \chi_1(0) - 1/2 - k \right|$, то виконується

$$\tilde{M}_n - \theta_n^2 \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \chi_2(0) - \frac{1}{12}.$$

Теорема 27. Нехай M_n – ширина випадкового рекурсивного дерева з $n + 1$ зовнішньою вершиною. З ймовірністю 1 множина границь підпоследовностей послідовності

$$\tilde{M}_n := 2 \log n \left(1 - \frac{\sqrt{2\pi \log n} M_n}{n} \right), \quad n \in \mathbb{N},$$

є інтервалом $[\chi_2(0) - 1/12, \chi_2(0) + 1/6]$. Більш того, якщо покласти $\theta_n := \min_{k \in \mathbb{Z}} \left| \log n + \chi_1(0) - 1/2 - k \right|$, то виконується

$$\tilde{M}_n - \theta_n^2 \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \chi_2(0) - \frac{1}{12}.$$

Теорема 28. Нехай M_n – ширина p -орієнтованого рекурсивного дерева з $pn + 1$ зовнішньою вершиною. З ймовірністю 1 множина границь підпоследовностей послідовності

$$\tilde{M}_n := \frac{2}{p} \log n \left(1 - \frac{\sqrt{2\pi \log n} M_n}{p^{3/2} n} \right), \quad n \in \mathbb{N},$$

є інтервалом $[\chi_2(0) - 1/12, \chi_2(0) + 1/6]$. Більш того, якщо покласти $\theta_n := \min_{k \in \mathbb{Z}} \left| \frac{1}{p} \log n + \chi_1(0) - 1/2 - k \right|$, то виконується

$$\tilde{M}_n - \theta_n^2 \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \chi_2(0) - \frac{1}{12}.$$

Доведення теорем 26, 27, 28. Ці результати є наслідками теорем 63 з додатку для однорозгалужених ГВБ, що відповідають випадковим деревам як описано в розділі 1.1.2, та леми 1.

Оскільки значення послідовності $(\text{frac}(a \log n))_{n \in \mathbb{N}}$ при $a > 0$ всюди щільні на $[0, 1]$, де $\text{frac}(x)$ позначає дробову частину x , то множина границь підпослідовностей послідовності θ_n^2 рівна $[0, \frac{1}{2}]$, звідки випливає результат про інтервал граничних точок послідовності \tilde{M}_n . \square

1.3 Математичне моделювання

Результати, що наведені в розділі 1.2, стверджують збіжність майже напевне, що означає, що відповідні асимптотичні властивості повинні виконуватись м.н. для будь-якої реалізації випадкового дерева. Хоча це результати суто теоретичної природи, тим не менш цікаво порівняти їх з реальними даними. Математичні симуляції, що представлені в цьому розділі, виконані на єдиному дереві, що збільшується з ростом n . Більш того, всі симуляції були проведені згідно з процедурою, що описана в розділі 1.1.2 (також більше детально дивись розділ 1.3.1).

Спочатку розглянемо 3-арне дерево. З теореми 7 при $r = 0$ ми отримуємо для всіх $k \in \mathbb{Z}$:

$$\mathbb{L}_n(k) - \mathbb{L}_n^*(k) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де

$$\mathbb{L}_n^*(k) = \frac{2n}{\sqrt{3\pi \log n}} e^{-\frac{(k - \frac{3}{2} \log n)^2}{3 \log n}}.$$

З теореми 19 маємо, що мода u_n при великих n має бути рівною $\lfloor u_n^* \rfloor$ чи $\lceil u_n^* \rceil$, де

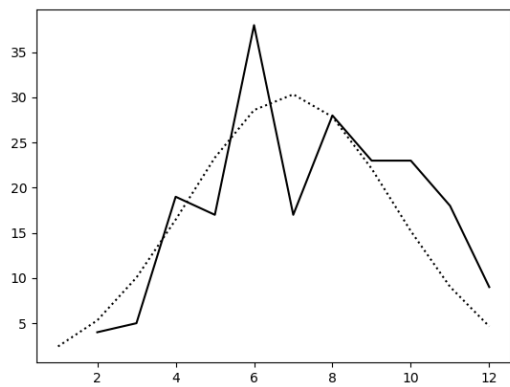
$$u_n^* = \frac{3}{2} \log n - \frac{1}{2}$$

та випадкова величина $\chi_1(0)$ прирівняна до 0 без підрахунку.

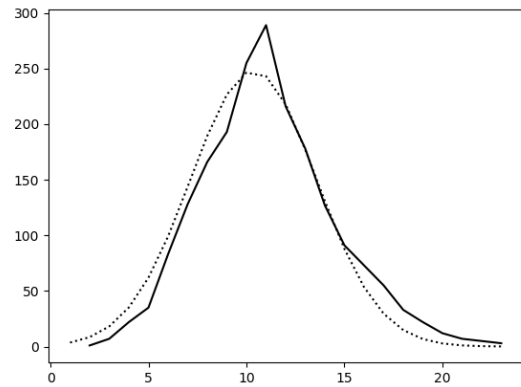
З теореми 23 отримуємо, що $M_n/M_n^* \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1$, де

$$M_n^* = \frac{2n}{\sqrt{3\pi \log n}}.$$

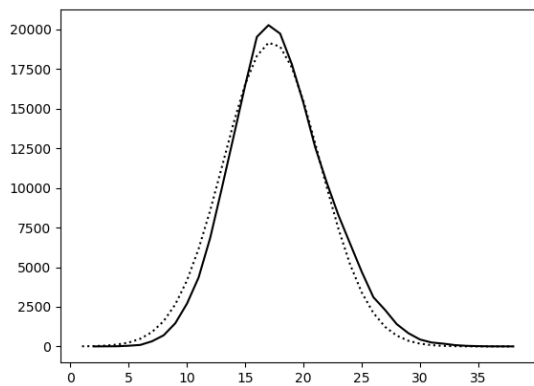
На рисунку 1.4 представлена еволюція профілю $\mathbb{L}_n(k)$ та його очікуване теоретичне значення $\mathbb{L}_n^*(k)$ змодельованого 3-арного дерева при різних n . В



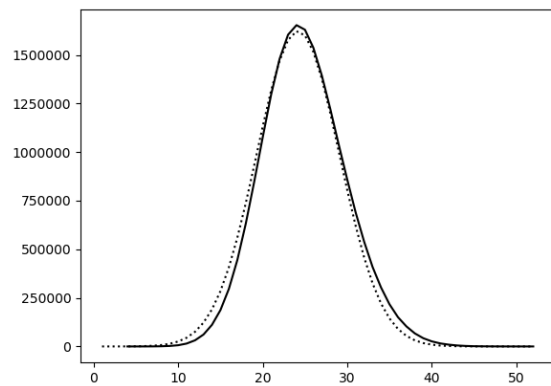
$n = 100$



$n = 1000$



$n = 100000$



$n = 10000000$

Рис. 1.4: Еволюція профілю $\mathbb{L}_n(k)$ 3-арного дерева (суцільна лінія) та його теоретичного значення $\mathbb{L}_n^*(k)$ (пунктирна лінія).

n	M_n	M_n^*	M_n/M_n^*	u_n	u_n^*
100	38	30.4	1.252	6	6.41
1000	289	247.9	1.166	11	9.86
10000	2511	2146.6	1.170	14	13.32
100000	20269	19200.0	1.056	17	16.77
1000000	179494	175271.4	1.024	21	20.22
10000000	1653079	1622697.5	1.019	24	23.68

Табл. 1.1: Еволюція ширини та моди 3-арного дерева.

таблиці 1.1 порівняно змодельовані та теоретичні значення моди та ширини цього дерева, коли n змінюється.

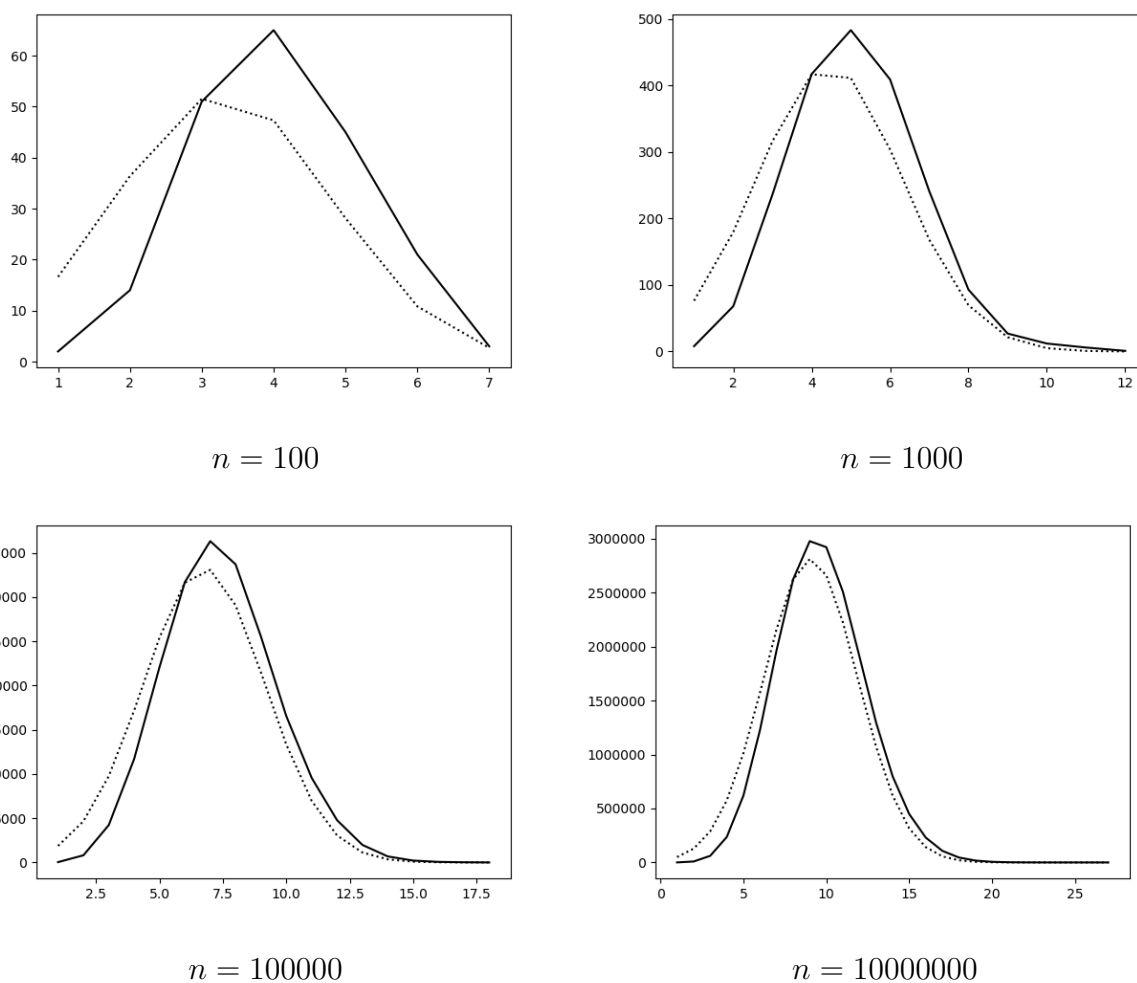


Рис. 1.5: Еволюція профілю $\mathbb{L}_n(k)$ ПОРДу (суцільна лінія) та його теоретичне значення $\mathbb{L}_n^*(k)$ (пунктирна лінія).

Продовжимо з моделювання ПОРДу. З результатів розділу 1.2 отримуємо

$$\mathbb{L}_n^*(k) = \frac{2n}{\sqrt{\pi \log n}} e^{-\frac{(k-1-\frac{1}{2} \log n)^2}{\log n}},$$

$$u_n^* = \frac{\log n}{2} + \frac{1}{2},$$

$$M_n^* = \frac{2n}{\sqrt{\pi \log n}}.$$

Результати моделювання представлені на рисунку 1.5 та в таблиці 1.2.

З вищенаведених симуляцій виглядає правдоподібно, що реальні значення збігаються до своїх теоретичних значень. Проте швидкість збіжності є досить повільною.

n	M_n	M_n^*	M_n/M_n^*	u_n	u_n^*
100	65	52.6	1.236	4	2.80
1000	483	429.3	1.125	5	3.95
10000	4114	3718.1	1.106	6	4.11
100000	36333	33255.4	1.093	7	6.26
1000000	327194	303578.9	1.078	8	7.41
10000000	2976165	2810594.5	1.059	9	8.56

Табл. 1.2: Еволюція ширини та моди ПОРДу.

1.3.1 Програмний код мовою Python3. В цьому підрозділі наведено програмний код для математичного моделювання, приклади результатів котрого були щойно викладені.

Для початку нам необхідно створити загальну модель для випадкових дерев з метою подальшого зручного використання й дослідження їх характеристик. Зауважимо, що оскільки досліджується зовнішній профіль, то моделювання можливо було б реалізувати легшим способом. Наприклад, бінарне дерево пошуку можна будувати на послідовності випадкових точок чи можна було б імітувати суто однорозгалуджене гіллясте випадкове блукання, що відповідає деревам. Проте, щоб уникнути можливих неточностей та заради більш доцільної перевірки теоретичних результатів, в наведеній реалізації уникнено всі спрощення та заміни, процедура побудови дерев відбувається в точності за схемами, що описані в підрозділі 1.1.2. Ці реалізації наведені в лістингу 1.1.

```

1 # sgrt.py
2 import random
3 import abc
4
5
6 class SimplyGeneratedRandomTree(abc.ABC):
7     # an abstract class for a simply generated random tree
8
9     def __init__(self):
10         self.graph = {} # oriented graph stored as an adjacency list
11         self.parent = {} # a parent of each note
12         self.external = [0] # list of all current external nodes

```

```

13     self._last_node = 0 # number of currently the last node in the graph
14     self.n = 0 # current iteration
15
16     def _get_next_node(self):
17         # getting a number for a new node to add
18         self._last_node += 1
19         return self._last_node
20
21     @abc.abstractmethod
22     def _add_subtree(self, v):
23         # adds a subtree to a chosen node v with respect to the point process
24         # returns a list of newly added external nodes
25         pass
26
27     @abc.abstractmethod
28     def _update_external_nodes(self, i, new_nodes):
29         # updates the list of all external nodes after adding new_nodes
30         # with a chosen node of index i
31         pass
32
33     def iterate(self, t=1):
34         # make t iterations of the simply generated tree
35         for _ in range(t):
36             i = random.randint(0, len(self.external) - 1)
37             new_nodes = self._add_subtree(self.external[i])
38             self._update_external_nodes(i, new_nodes)
39             self.n += 1
40
41     def get_profile(self):
42         # returns an external profile  $L_n(k)$ 
43         profile = {}
44         external_nodes = set(self.external)
45
46         # deep first search
47         def dfs(v, level):
48             if v in external_nodes:
49                 profile[level] = profile.get(level, 0) + 1
50                 for w in self.graph.get(v, []):
51                     dfs(w, level + 1)
52
53         dfs(0, 0)
54         return profile
55
56     def get_width_and_mode(self, profile=None):
57         # returns a width  $M_n$  and a mode  $u_n$  simultaneously

```

```

58     # if there are many modes, returns any of them
59     if profile is None:
60         profile = self.get_profile()
61         max_width = -1
62         mode = None
63         for k in profile:
64             if profile[k] > max_width:
65                 max_width = profile[k]
66                 mode = k
67         return max_width, mode
68
69
70 class RandomRecursiveTree(SimplyGeneratedRandomTree):
71     # a class for imitating a random recursive tree
72
73     def __init__(self):
74         super().__init__()
75
76     def _add_subtree(self, v):
77         # adds a subtree to a chosen node v with respect to the point process
78         # returns a list of newly added external nodes
79         new_node = self._get_next_node()
80         self.graph[v] = self.graph.get(v, []) + [new_node]
81         self.parent[new_node] = v
82         return new_node,
83
84     def _update_external_nodes(self, i, new_nodes):
85         # updates the list of all external nodes after adding new_nodes
86         # with a chosen node of index i
87         self.external.append(new_nodes[0])
88
89
90 class DaryTree(SimplyGeneratedRandomTree):
91     # a class for imitating a D-ary tree
92
93     def __init__(self, D):
94         # D >= 1
95         super().__init__()
96         self.D = D
97
98     def _add_subtree(self, v):
99         new_nodes = [self._get_next_node() for _ in range(self.D)]
100        self.graph[v] = new_nodes
101        for new_node in new_nodes:
102            self.parent[new_node] = v

```

```

103     return tuple(new_nodes)
104
105     def _update_external_nodes(self, i, new_nodes):
106         self.external[i] = new_nodes[0]
107         # replacing a chosen node, making it internal
108         for new_node in new_nodes[1:]:
109             self.external.append(new_node)
110
111
112 class POrientedRecursiveTree(SimplyGeneratedRandomTree):
113     # a class for imitating a p-oriented recursive tree
114
115     def __init__(self, p):
116         # p >= 1
117         super().__init__()
118         self.p = p
119         # PORTs have different start
120         self.graph = {0: [1]}
121         self.parent = {1: 0}
122         self.external = [1]
123         self._last_node = 1
124
125     def _add_subtree(self, v):
126         new_nodes = [self._get_next_node() for _ in range(self.p)]
127         # adding p new nodes on the same level as v
128         self.graph[self.parent[v]] += new_nodes
129         for new_node in new_nodes:
130             self.parent[new_node] = self.parent[v]
131         new_node = self._get_next_node() # adding a new node as a son of v
132         self.graph[v] = [new_node]
133         self.parent[new_node] = v
134         return tuple(new_nodes + [new_node])
135
136     def _update_external_nodes(self, i, new_nodes):
137         self.external[i] = new_nodes[0]
138         # replacing a chosen node, making it internal
139         for new_node in new_nodes[1:]:
140             self.external.append(new_node)

```

Лістинг 1.1: Реалізації випадкових дерев простої будови.

Абстрактний клас `SimplyGeneratedRandomTree` описує загальну модель випадкового дерева простої будови. Сам граф описується за допомогою списку суміжності, також додатково зберігаються список усіх поточних зовні-

шніх вершин та батьків кожної вершини. На початку граф складається з лише однієї зовнішньої вершини. Кожна ітерація побудови дерева (метод `iterate()`) складається з двох кроків: додавання нового піддерева до випадково обраної зовнішньої вершини та оновлення списку зовнішніх вершин. В першу чергу саме ці дві операції і задають різні типи дерев та мають бути описані в дочірніх класах. Функція `get_profile()` підраховує поточний профіль дерева за допомогою його обходу в глибину, функція `get_width_and_mode()` водночас підраховує ширину та моду дерева, оскільки функціонали визначені як максимум та де він досягається профілю дерева. Наступні три дочірні класи задають відповідні типи дерев, що були досліджені в цьому розділі.

У лістингу 1.2 наведено приклад використання вищеописаних класів для генерування 3-арного дерева та порівняння його характеристик з теоретичними. Результати його виконання було наведено на рисунку 1.4 та в таблиці 1.1

```

1 from math import log, exp, sqrt, pi
2 import matplotlib.pyplot as plt
3 import sgrt
4
5
6 def get_theoretical_profile(D, n, k):
7     L = n * sqrt((D - 1)**3) / sqrt(2 * D * pi * log(n)) * \
8         exp(-(k - D / (D - 1) * log(n))**2 / (2 * D / (D - 1) * log(n)))
9     return L
10
11
12 def get_theoretical_width(D, n):
13     width = n * sqrt((D - 1) ** 3) / (sqrt(2 * D * pi * log(n)))
14     return width
15
16
17 def get_theoretical_mode(D, n):
18     mode = D / (D - 1) * log(n) - 0.5
19     return mode
20
21
22 # checking tree's characteristics on each power of 10
23 # making up to 10^7 iterations
24 N = [10**i for i in range(2, 8)]

```

```

25
26 D = 3
27 tree = sgrt.DaryTree(D=D)
28
29 for t in range(len(N)):
30     if t == 0:
31         n = N[0]
32     else:
33         # what is left to iterate considering previous iterations
34         n = N[t] - N[t - 1]
35     tree.iterate(n)
36
37     profile = tree.get_profile()
38     width, mode = tree.get_width_and_mode(profile)
39
40     theoretical_width = get_theoretical_width(D, N[t])
41     theoretical_mode = get_theoretical_mode(D, N[t])
42
43     # printing a table
44     print(N[t], width, theoretical_width, width/theoretical_width,
45           mode, theoretical_mode)
46
47     # making a plot for profile
48     fig, ax = plt.subplots()
49
50     x = []
51     y = []
52     for k in sorted(list(profile.keys())):
53         x.append(k)
54         y.append(profile[k])
55     ax.plot(x, y, "-", color="black")
56
57     x = []
58     y = []
59     for k in range(1, max(profile.keys()) + 1):
60         x.append(k)
61         y.append(get_theoretical_profile(D, N[t], k))
62     ax.plot(x, y, ":", color="black")

```

Лістинг 1.2: Код для генерування 3-арного дерева та порівняння його характеристик з теоретичними.

Аналогічно у лістингу 1.3 наведена реалізація ПОРДу для виведення його змодельованих та теоретичних характеристик у вигляді таблиці та графіків,

результати чого можуть бути знайдені в підрозділі 1.3. Зауважимо, що реалізація класу POrientedRecursiveTree трохи відрізняється, оскільки p -орієнтовані рекурсивні дерева мають унікальну структуру на нульовій ітерації, що також знаходить відображення й у зміщених теоретичних значеннях.

```

1  from math import log, exp, sqrt, pi
2  import matplotlib.pyplot as plt
3  import sgrt
4
5
6  def get_theoretical_profile(p, n, k):
7      k -= 1
8      L = n * sqrt(p**3) / sqrt(2 * pi * log(n)) * \
9          exp(-p * (k - log(n) / p)**2 / (2 * log(n)))
10     return L
11
12
13 def get_theoretical_width(p, n):
14     width = n * sqrt(p**3) / sqrt(2 * pi * log(n))
15     return width
16
17
18 def get_theoretical_mode(p, n):
19     mode = log(n) / p - 0.5 + 1
20     return mode
21
22
23 # checking tree's characteristics on each power of 10
24 # making up to 10^7 iterations
25 N = [10**i for i in range(2, 8)]
26
27 p = 2
28 tree = sgrt.POrientedRecursiveTree(p=p)
29
30 for t in range(len(N)):
31     if t == 0:
32         n = N[0]
33     else:
34         # what is left to iterate considering previous iterations
35         n = N[t] - N[t - 1]
36     tree.iterate(n)
37
38     profile = tree.get_profile()
39     width, mode = tree.get_width_and_mode(profile)

```

```

40
41     theoretical_width = get_theoretical_width(p, N[t])
42     theoretical_mode = get_theoretical_mode(p, N[t])
43
44     # printing a table
45     print(N[t], width, theoretical_width, width/theoretical_width,
46           mode, theoretical_mode)
47
48     # making a plot for profile
49     fig, ax = plt.subplots()
50
51     x = []
52     y = []
53     for k in sorted(list(profile.keys())):
54         x.append(k)
55         y.append(profile[k])
56     ax.plot(x, y, "-", color="black")
57
58     x = []
59     y = []
60     for k in range(1, max(profile.keys()) + 1):
61         x.append(k)
62         y.append(get_theoretical_profile(p, N[t], k))
63     ax.plot(x, y, ":", color="black")

```

Лістинг 1.3: Код для генерування ПОРДу та порівняння його характеристик з теоретичними.

1.4 Висновки до розділу 1

В цьому розділі дисертаційної роботи були досліджені профілі випадкових рекурсивних дерев, D -арних дерев та p -орієнтованих рекурсивних дерев. Зокрема, були отримані наступні результати щодо профілів:

- Отримані асимптотичні розклади майже напевне для профілів вищенаведених типів випадкових дерев з довільною кількістю членів.
- Наведені результати щодо асимптотичної поведінки профілів, у випадках, коли досліджуваний рівень зростає разом з кількістю вершин у дереві.

- Отримані результати щодо пов'язаних з профілями характеристик дерев, а сами отримані збіжності майже напевне для моди та ширини дерев.

Отримані твердження частково повторюють вже відомі результати для певних класів дерев, що були доведені раніше іншими методами та в меншій загальності.

Для випадкових величин, що фігурують в асимптотичних розкладах, були знайдені нові характеристики у вигляді стохастичних рівнянь нерухомої точки.

Додатково основні результати щодо профілів, моди та ширини було перевірено за допомогою математичного моделювання.

Серед напрямків подальшого дослідження можна виділити наступні:

- Оскільки техніка, що використовувалась для отримання вищенаведених результатів, є досить загальною й може бути застосована й для інших випадкових дерев простої будови, якщо виникне така необхідність, це наводить на думку, що питання щодо асимптотичної поведінки профілів можуть бути закриті. Проте, відкритим залишається питання про асимптотичні властивості профілів дерев простої будови, що не задовольняють припущенням леми 1.
- Цікавим залишається подальше дослідження випадкових величин, що фігурують в асимптотичних розкладах, а саме явне знаходження цих величин, їх функцій розподілу тощо, в тому числі за допомогою отриманих в цьому розділі роботи характеристик цих величин.

Розділ 2

Дерева загальних гіллястих процесів

2.1 Означення та обговорення

Нехай $(\xi_i, \eta_i)_{i \in \mathbb{N}}$ – послідовність незалежних копій випадкового вектора (ξ, η) з майже напевно додатними компонентами, що можуть бути довільно залежними. Означимо через $S := (S_i)_{i \geq 0}$ звичайне випадкове блукання з кроками ξ_i для $i \in \mathbb{N}$, тобто $S_0 := 0$ та $S_i := \xi_1 + \dots + \xi_i$ при $i \in \mathbb{N}$. Покладемо

$$T_i := S_{i-1} + \eta_i, \quad i \in \mathbb{N}.$$

Послідовність $T := (T_i)_{i \in \mathbb{N}}$ називається *збуреним випадковим блуканням*.

Тепер наведемо конструкцію загального гіллястого процесу (чи гіллястого процесу Крампа-Мода-Ягерса) в окремому випадку, коли він згенерований за допомогою T . В момент часу 0 присутня єдина особа. Вона породжує потомство (перше покоління), часи народження котрого задані точками T . Перше покоління породжує друге покоління. Зсуви часів народження осіб наступного покоління відносно часу народження їх матері розподілені як копії T , й для різних батьків ці копії незалежні.

На рисунку 2.1 зображено певну реалізацію такого загального гіллястого випадкового блукання. Верхній індекс в $T_r^{(j)}$ позначає номер покоління, а нижній визначає номер особи (не обов'язково пронумерований зліва направо). Зсуви часів народження кожного покоління відносно матері розподілені як

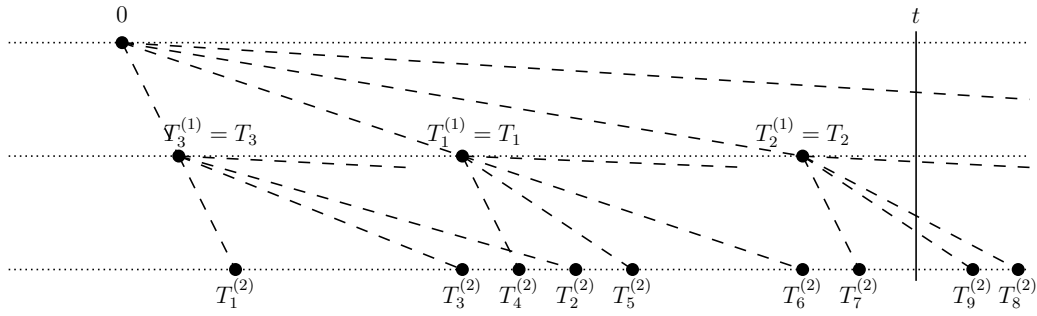


Рис. 2.1: Приклад загального гіллястого процесу згенерованого за T .

незалежна копія T , тобто тут $T_7^{(2)} - T_2^{(1)}$, $T_9^{(2)} - T_2^{(1)}$ та $T_8^{(2)} - T_2^{(1)}$ розподілені як перші три найменші елементи T . Зауважимо також, що в загальному послідовність $T = (T_k)_{k \in \mathbb{N}}$ (в порівнянні з елементами S) не обов'язково монотонна.

На рисунку також наведено деяке порогове значення t , якщо залишити лише осіб з часом народження, що не перевищує t , то дана конструкція реалізує так зване дерево загального гіллястого процесу в момент t , для якого можуть бути досліджені класичні характеристики такі як, наприклад, висота.

Очевидно, що випадкова послідовність $T^{(j)}$ ($j \geq 2$), що визначена часами народження осіб в j -му поколінні загального гіллястого процесу, більш складна ніж збурене випадкове блукання $T^{(1)} = T$. Називатимемо $(T^{(j)})_{j \geq 2}$ ітерованим збуреним випадковим блуканням на дереві загального гіллястого процесу.

Для $t \geq 0$ та $j \in \mathbb{N}$, позначимо через $N_j(t)$ кількість осіб в j -му поколінні з часами народження не більшими за t . Наприклад, на рисунку 2.1 $N_1(t) = N(t) = 3$ та $N_2(t) = 7$. Для $r \in \mathbb{N}$ нехай $N_{j-1}^{(r)}(t)$ визначає число нащадків в j -му поколінні з часами народження з інтервалу $[T_r, t + T_r]$ особи першого покоління з часом народження T_r . За властивістю розгалуження $(N_{j-1}^{(1)}(t))_{t \geq 0}$, $(N_{j-1}^{(2)}(t))_{t \geq 0}$, \dots незалежні копії N_{j-1} , котрі також незалежні від T . Простий розклад, що демонструє властивості процесів $N_j := (N_j(t))_{t \geq 0}$ й також демонструє її рекурсивну структуру, такий

$$N_j(t) = \sum_{r \geq 1} N_{j-1}^{(r)}(t - T_r) \mathbb{1}_{\{T_r \leq t\}}, \quad j \geq 2, \quad t \geq 0.$$

Більш того, нехай $T^{(j-1)} := (T_r^{(j-1)})_{r \geq 1}$ – деяка нумерація часів народження осіб $(j-1)$ -го покоління; $N_{1,j}^{(r)}(t)$ число дітей в j -у покоління з часами народження в інтервалі $[T_r^{(j-1)}, t + T_r^{(j-1)}]$ особи $(j-1)$ -го покоління з часом народження $T_r^{(j-1)}$. Знову, за властивістю розгалуження, $(N_{1,j}^{(1)}(t))_{t \geq 0}$, $(N_{1,j}^{(2)}(t))_{t \geq 0}, \dots$ незалежні копії $(N(t))_{t \geq 0}$, які також не залежать від $T^{(j-1)}$. Таким чином, можемо записати рекурсивний розклад N_j :

$$N_j(t) = \sum_{r \geq 1} N_{1,j}^{(r)}(t - T_r^{(j-1)}) \mathbb{1}_{\{T_r^{(j-1)} \leq t\}}, \quad j \geq 2, \quad t \geq 0.$$

Тепер розглянемо частковий випадок, коли $\eta = \xi$ м.н., тоді $T = S$ м.н., та відповідна послідовність $(S^{(j)})_{j \geq 2}$ може називатись *ітерованим випадковим блуканням на дереві загального гіллястого процесу*. Це поняття не слід плутати з ітерованими процесами відновлення, що розглядалися, наприклад, в [71].

В цьому випадку $N_1(t) = N(t) = \sum_{i \geq 1} \mathbb{1}_{\{T_i \leq t\}} = \sum_{i \geq 1} \mathbb{1}_{\{S_i \leq t\}}$ є класичним процесом відновлення. Позначимо через $V_1(t) = \mathbb{E}N(t)$ середню кількість осіб в першому поколінні, що народились до моменту часу t включно. Значення останнього функціонала для наступних поколінь ($j \geq 2$) можна порахувати за допомогою згортки¹. Очевидна наступна рекурсивна залежність:

$$V_j(t) = (V_{j-1} * V)(t) = \int_{[0,t]} V_{j-1}(t-y) dV(y), \quad j \geq 2, \quad t \geq 0.$$

Тут і надалі $u * v$ позначає згортку Лебега-Стілтєса двох функцій u та v локально обмеженої варіації. Також позначатимемо через $u^{*(j)}$, $j \in \mathbb{N}$, j -ту кратну згортку функції u .

Також зауважимо, що у випадку $\eta = \xi$ $V_1(t) = U(t) - 1$ при $t \geq 0$, де $U(t) = \mathbb{E} \inf\{i \in \mathbb{N} : S_i > t\}$ – звичайна функція відновлення. Нагадаємо, що функція відновлення не спадає, неперервна справа та $U(t) = 0$ при $t < 0$ (див., наприклад, [2]). Властивості цих згорток для функції відновлення будуть зазначені у наступних підрозділах, у випадках, коли j не є константою та зростає до нескінченності з ростом t .

¹Цей підхід справедливий і для довільного вибору η .

Зауважимо, що у випадку, коли ξ має експоненційний розподіл з параметром 1, зображення на рисунку 2.1 з обмеженням t фактично є реалізацією випадкового рекурсивного дерева.

2.2 Результати для кратних згорток функцій лінійного росту

Наведені тут результати в першу чергу стосуються наступної суто аналітичної проблеми. Припустимо, що неспадна функція f зростає лінійно, тобто $f(t) \sim at$ при $t \rightarrow \infty$ для деякої $a > 0$. Тоді, для фіксованого $j \in \mathbb{N}$,

$$f^{*(j)}(t) \sim \frac{a^j t^j}{j!}, \quad t \rightarrow \infty.$$

Зробивши різні припущення відносно поведінки $f(t) - at$, ми розширимо цю асимптотику на випадок, коли $j = j(t)$ розбігається до нескінченності при $t \rightarrow \infty$. Відповідні твердження застосовуються для дослідження асимптотики згорток функцій відновлення.

Теорема 29. *Нехай $f : \mathbb{R} \rightarrow [0, \infty)$ позначає неспадну неперервну справа функцію, що обертається в нуль на від'ємній півпрямій, та задовольняє*

$$f(t) = at + O(t^\alpha), \quad t \rightarrow \infty \tag{2.1}$$

для деяких $a > 0$ та $\alpha \in [0, 1)$. Тоді, для будь-якої цілозначної функції $j = j(t)$ такої, що $j(t) = o(t^{(1-\alpha)/2})$ при $t \rightarrow \infty$, виконується

$$f_j(t) := f^{*(j)}(t) \sim \frac{a^j t^j}{j!}, \quad t \rightarrow \infty.$$

Доведення. Згідно з припущенням (2.1) існує константа $C \geq 1$ така, що

$$-C(t+1)^\alpha \leq f(t) - at \leq C(t+1)^\alpha, \quad t \geq 0. \tag{2.2}$$

Для $j \in \mathbb{N}$ та $t \geq 0$ покладемо

$$r_j(t) := \int_{[0,t]} f_j(t-y) d(f(y) - ay) = \int_{[0,t]} (f(t-y) - a(t-y)) df_j(y)$$

та зауважимо, що

$$f_j(t) = r_{j-1}(t) + a \int_0^t f_{j-1}(y) dy, \quad j \geq 2, \quad t \geq 0.$$

В силу (2.2), робимо висновок, що

$$|r_j(t)| \leq C(t+1)^\alpha f_j(t), \quad j \in \mathbb{N}, \quad t \geq 0.$$

Використовуючи цю оцінку та математичну індукцію, отримуємо

$$W_j^-(t) \leq f_j(t) \leq W_j^+(t), \quad j \in \mathbb{N}, \quad t \geq 0, \quad (2.3)$$

де W_j^\pm визначається рекурсивно формулами $W_0^\pm(t) := 1$ та

$$W_j^\pm(t) = \left(\pm C(t+1)^\alpha W_{j-1}^\pm(t) + a \int_0^t W_{j-1}^\pm(y) dy \right)_+, \quad j \in \mathbb{N}, \quad t \geq 0.$$

Тут нагадуємо, що $x_+ = \max(x, 0)$, та зауважимо, що взяття невід'ємної частини важливо лише для W_j^- , що забезпечує її невід'ємність, тоді як цю операцію можна пропустити для W_j^+ .

Залишається показати, що

$$W_j^\pm(t) \sim \frac{a^j t^j}{j!}, \quad t \rightarrow \infty. \quad (2.4)$$

Для цього спочатку доведемо методом математичної індукції, що

$$W_j^+(t) \leq \frac{a^j t^j}{j!} + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!}, \quad j \in \mathbb{N}, \quad t \geq 0. \quad (2.5)$$

При $j = 1$ ця нерівність виконується тривіально, оскільки $W_1^+(t) = C(t+1)^\alpha + at$. Крок індукції перевіряється так

$$\begin{aligned} W_{j+1}^+(t) &\leq C(t+1)^\alpha \left(\frac{a^j t^j}{j!} + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!} \right) \\ &\quad + a \int_0^t \left(\frac{a^j y^j}{j!} + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (y+1)^{\alpha(j-i)+i}}{i!} \right) dy \\ &\leq \sum_{i=0}^j \binom{j}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!} + \frac{a^{j+1} t^{j+1}}{(j+1)!} \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^{i+1} C^{j-i}}{i!} \frac{(t+1)^{\alpha(j-i)+i+1}}{\alpha(j-i)+i+1} \\
& \leq \frac{a^{j+1} t^{j+1}}{(j+1)!} + \sum_{i=0}^j \binom{j}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!} \\
& \quad + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^{i+1} C^{j-i}}{(i+1)!} (t+1)^{\alpha(j-i)+i+1} \\
& = \frac{a^{j+1} t^{j+1}}{(j+1)!} + \sum_{i=0}^j \binom{j}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!} \\
& \quad + \sum_{i=1}^j \binom{j}{i-1} \frac{a^i C^{j+1-i}}{i!} (t+1)^{\alpha(j+1-i)+i} \\
& = \frac{a^{j+1} t^{j+1}}{(j+1)!} + \sum_{i=0}^j \binom{j+1}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!},
\end{aligned}$$

де в останньому переході була використана біноміальна тотожність $\binom{j}{i} + \binom{j}{i-1} = \binom{j+1}{i}$. Більш того, використовуючи $j(t) = o(t)$ отримуємо

$$\begin{aligned}
\frac{j!}{a^j t^j} \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!} & \sim \frac{j!}{a^j (t+1)^j} \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!} \\
& \leq \sum_{i=0}^{j-1} \left(\frac{j!}{i!} \right)^2 \left(\frac{C}{a} \right)^{j-i} (t+1)^{(1-\alpha)(i-j)} \\
& \leq \sum_{i=0}^{j-1} (j^{j-i})^2 (Ca^{-1})^{j-i} (t+1)^{(1-\alpha)(i-j)} \\
& \leq \sum_{i \geq 1} \left(\frac{Ca^{-1} j^2}{(t+1)^{1-\alpha}} \right)^i \\
& = \frac{Ca^{-1} j^2}{(t+1)^{1-\alpha}} \left(1 - \frac{Ca^{-1} j^2}{(t+1)^{1-\alpha}} \right)^{-1}.
\end{aligned}$$

Отже, з огляду на припущення $j(t) = o(t^{(1-\alpha)/2})$ маємо

$$\limsup_{t \rightarrow \infty} \frac{j!}{a^j t^j} W_j^+(t) \leq 1. \quad (2.6)$$

Щоб довести

$$W_j^-(t) \geq \left(\frac{a^j t^j}{j!} - \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!} \right)_+, \quad j \in \mathbb{N}, \quad t \geq 0 \quad (2.7)$$

ми використаємо подібні міркування. Формула (2.7) очевидно виконується при $j = 1$, й виводимо її за допомогою математичної індукції для $j \geq 2$ так:

$$\begin{aligned}
W_{j+1}^-(t) &\geq -C(t+1)^\alpha W_j^-(t) + a \int_0^t W_j^-(y) dy \\
&\geq -C(t+1)^\alpha W_j^+(t) + a \int_0^t W_j^-(y) dy \\
&\geq -C(t+1)^\alpha W_j^+(t) \\
&\quad + a \int_0^t \left(\frac{a^j y^j}{j!} - \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (y+1)^{\alpha(j-i)+i}}{i!} \right) dy \\
&\geq -C(t+1)^\alpha \left(\frac{a^j t^j}{j!} + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (t+1)^{\alpha(j-i)+i}}{i!} \right) \\
&\quad + a \int_0^t \left(\frac{a^j y^j}{j!} - \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^i C^{j-i} (y+1)^{\alpha(j-i)+i}}{i!} \right) dy \\
&\geq \frac{a^{j+1} t^{j+1}}{(j+1)!} - \left(\sum_{i=0}^j \binom{j}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!} \right. \\
&\quad \left. + \sum_{i=0}^{j-1} \binom{j}{i} \frac{a^{i+1} C^{j-i}}{i!} \int_0^t (y+1)^{\alpha(j-i)+i} dy \right) \\
&\geq \frac{a^{j+1} t^{j+1}}{(j+1)!} - \sum_{i=0}^j \binom{j+1}{i} \frac{a^i C^{j+1-i} (t+1)^{\alpha(j+1-i)+i}}{i!}.
\end{aligned}$$

Ми використали (2.3) та (2.5) в другій та четвертій нерівностях, відповідно. Оскільки W_{j+1}^- невід'ємна, то отримуємо (2.7). Таким чином

$$\liminf_{t \rightarrow \infty} \frac{j!}{a^j t^j} W_j^-(t) \geq 1. \tag{2.8}$$

Поєднання (2.6) та (2.8) призводить до (2.4), що завершує доведення теореми 29. □

Наслідок 30. Нехай $\mathbb{E}\xi^2 < \infty$, тоді

$$U^{*(j)}(t) \sim \frac{t^j}{\mu^j j!}$$

при $t \rightarrow \infty$, де $\mu = \mathbb{E}\xi$ та $j = j(t)$ – цілочисельна функція, для якої $j(t) = o(t^{1/2})$ при $t \rightarrow \infty$.

Доведення. Відомо ([2, твердження 6.2, с. 160]), що за умови скінченності другого (й, відповідно, першого) моменту ξ функція відновлення має наступну асимптотичну поведінку:

$$U(t) = \frac{t}{\mu} + O(1) \quad (2.9)$$

при $t \rightarrow \infty$. Тоді твердження наслідку безпосередньо випливає з теореми 29. \square

Цей результат за тих самих умов на в.в. ξ вже було доведено в роботі [14], проте менш загальним підходом.

У наступному результаті наведено асимптотики кратних згортки $f^{*(j)}$ при $j = j(t)$, що можуть зростати швидше ніж $t^{1/2}$, за припущень, що функція $|f(t) - at|$ має скінченну повну варіацію та задовольняє додаткове припущення щодо інтегрованості. Ми використаємо домовленість, що для функції $x : \mathbb{R} \rightarrow \mathbb{R}$, $x^{*(0)}(t) = \mathbb{1}_{[0, \infty)}(t)$, $t \in \mathbb{R}$. Також позначатимемо через $\mathcal{V}_I(x)$ повну варіацію x на (можливо нескінченному) інтервалі I . Нарешті, якщо x – функція зі скінченною повною варіацією на $[a, b]$, $-\infty \leq a < b \leq \infty$ та y є вимірною функцією на I , визначимо

$$\int_{[a, b]} y(t) |dx(t)| = \int_{[a, b]} y(t) d(\mathcal{V}_{[a, t]}(x)),$$

де інтеграл в правій частині розуміється в сенсі Лебега-Стільтєса.

Теорема 31. *Нехай $f : \mathbb{R} \mapsto [0, \infty)$ позначає неспадну неперервну справа функцію, що обертається в нуль на від'ємній півпрямій. Припустимо, що функція ε , визначена як*

$$\varepsilon(t) := f(t) - at, \quad t \geq 0, \quad (2.10)$$

для деякого $a > 0$, задовольняє

$$\int_{[0, \infty)} y |d\varepsilon(y)| < \infty. \quad (2.11)$$

Тоді, для будь-якої цілочисельної функції $j = j(t)$ такої, що $j(t) = o(t^{2/3})$ при $t \rightarrow \infty$, виконується

$$f_j(t) := f^{*(j)}(t) \sim \frac{a^j t^j}{j!} \exp\left(\frac{\gamma_0 j^2}{at}\right), \quad t \rightarrow \infty, \quad (2.12)$$

$$\partial e \gamma_0 := \int_{[0, \infty)} d\varepsilon(y) = \lim_{t \rightarrow \infty} (f(t) - at).$$

Доведення. Функція ε як різниця двох неспадних функцій має скінченну повну варіацію на будь-якому скінченному інтервалі. Зокрема, (2.11) тягне за собою

$$\int_{[0, \infty)} |d\varepsilon(y)| \leq \int_{[0, 1)} |d\varepsilon(y)| + \int_{[1, \infty)} y |d\varepsilon(y)| < \infty.$$

Тому ε має скінченну повну варіацію на $[0, \infty)$. Запишемо

$$\begin{aligned} \int_0^\infty |\varepsilon(y) - \gamma_0| dy &= \int_0^\infty \left| \int_{(y, \infty)} d\varepsilon(z) \right| dy \leq \int_0^\infty \int_{(y, \infty)} |d\varepsilon(z)| dy \\ &= \int_{[0, \infty)} y |d\varepsilon(y)| < \infty, \end{aligned}$$

де ми використали інтегрування частинами в останній рівності. Отже, (2.11) означає, що

$$\int_0^\infty |\varepsilon(y) - \gamma_0| dy < \infty. \quad (2.13)$$

Тепер змінимо (2.10) в околі нуля так, що основні властивості ε , що наведені в (2.11) та (2.13), будуть збережені. Покладемо

$$f(t) = (at + \gamma_0)_+ + \tilde{\varepsilon}(t) =: \ell(t) + \tilde{\varepsilon}(t), \quad t \in \mathbb{R}. \quad (2.14)$$

Зауважимо, що обидва доданки можуть бути ненульовими в обмеженому лівому околі нуля, все ж

$$\int_{\mathbb{R}} |\tilde{\varepsilon}(y)| dy < \infty \quad \text{та} \quad \int_{\mathbb{R}} |y| |d\tilde{\varepsilon}(y)| < \infty, \quad (2.15)$$

оскільки $t \mapsto \varepsilon(t) - \gamma_0 - \tilde{\varepsilon}(t)$ має обмежений носій. Перевага (2.14) обумовлена двома фактами:

(i) проста формула кратної згортки для ℓ , а саме

$$\ell^{*(j)}(t) = \frac{(at + \gamma_0 j)_+^j}{j!}, \quad j \in \mathbb{N}, \quad t \in \mathbb{R}, \quad (2.16)$$

(ii) функція $\tilde{\varepsilon}$ згасає достатньо швидко та, як така, є асимптотично незначною в сенсі, що $\ell^{*(j)}(t) \sim f^{*(j)}(t)$ при $t \rightarrow \infty$.

Щоб перевірити (2.16) використаємо математичну індукцію. При $j = 1$ формула тривіальна. Крок математичної індукції робиться наступним чином: для $t \geq -a^{-1}\gamma_0(j+1)$,

$$\begin{aligned}\ell^{*(j+1)}(t) &= \int_{\mathbb{R}} \frac{(a(t-y) + \gamma_0 j)_+^j}{j!} d\ell(y) \\ &= a \int_{-\gamma_0 a^{-1}}^{t+j\gamma_0 a^{-1}} \frac{(a(t-y) + \gamma_0 j)^j}{j!} dy \\ &= \int_0^{at+\gamma_0(j+1)} \frac{z^j}{j!} dz \\ &= \frac{(at + \gamma_0(j+1))^{j+1}}{(j+1)!},\end{aligned}$$

та $\ell^{*(j+1)}(t) = 0$ при $t < -a^{-1}\gamma_0(j+1)$.

Що стосується пункту (ii), використовуючи (2.14), комутативність та дистрибутивність згортки Лебега-Стілтєса, отримуємо

$$f^{*(j)}(t) = \ell^{*(j)}(t) + \sum_{k=0}^{j-1} \binom{j}{k} \left(\ell^{*(k)} * \tilde{\varepsilon}^{*(j-k)} \right) (t), \quad t \in \mathbb{R}.$$

Нижче буде показано, що другим доданком можна асимптотично знехтувати, коли $j(t) = o(t^{2/3})$. Припустимо, що це вже було показано. Тоді (2.12) впливає миттєво, бо, для достатньо великих t ,

$$f^{*(j)}(t) \sim \ell^{*(j)}(t) = \frac{a^j t^j}{j!} \left(1 + \frac{\gamma_0 j}{at} \right)^j = \frac{a^j t^j}{j!} \exp \left(j \log \left(1 + \frac{\gamma_0 j}{at} \right) \right).$$

Права частина асимптотично еквівалентна до $\frac{a^j t^j}{j!} \exp \left(\frac{\gamma_0 j^2}{at} \right)$, коли $j = j(t) = o(t^{2/3})$ при $t \rightarrow \infty$.

Повертаючись до аналізу

$$R_j(t) := \sum_{k=0}^{j-1} \binom{j}{k} \left(\ell^{*(k)} * \tilde{\varepsilon}^{*(j-k)} \right) (t), \quad t \geq 0,$$

спочатку перевіримо, що

$$\mathcal{V}_{\mathbb{R}}(\ell * \tilde{\varepsilon}) \leq \tilde{C} < \infty \tag{2.17}$$

для деякої абсолютної константи $\tilde{C} > 0$. Для $t \in \mathbb{R}$

$$(\ell * \tilde{\varepsilon})(t) = \int_{\mathbb{R}} \tilde{\varepsilon}(t-y) d\ell(y) = a \int_{-a^{-1}\gamma_0}^{\infty} \tilde{\varepsilon}(t-y) dy = a \int_{-\infty}^{t+a^{-1}\gamma_0} \tilde{\varepsilon}(y) dy.$$

Таким чином (2.17) виконується з $\tilde{C} := a \int_{\mathbb{R}} |\tilde{\varepsilon}(y)| dy$. Покладемо

$$g_{i,j}(t) := \mathcal{V}_{(-\infty, t]}(\ell^{*(i)} * \tilde{\varepsilon}^{*(j)}), \quad i, j \in \mathbb{N}_0, \quad t \in \mathbb{R}.$$

Тоді, для $i, j \in \mathbb{N}$, взявши до уваги комутативність згортки Лебега-Стілтєса, отримуємо

$$\begin{aligned} g_{i,j}(t) &= \mathcal{V}_{(-\infty, t]} \left((\ell^{*(i-1)} * \tilde{\varepsilon}^{*(j-1)}) * (\ell * \tilde{\varepsilon}) \right) \\ &\leq \mathcal{V}_{(-\infty, t]}(\ell^{*(i-1)} * \tilde{\varepsilon}^{*(j-1)}) \mathcal{V}_{(-\infty, t]}(\ell * \tilde{\varepsilon}) \\ &\leq \mathcal{V}_{(-\infty, t]}(\ell^{*(i-1)} * \tilde{\varepsilon}^{*(j-1)}) \mathcal{V}_{\mathbb{R}}(\ell * \tilde{\varepsilon}) \\ &\leq \tilde{C} g_{i-1, j-1}(t), \quad t \in \mathbb{R}. \end{aligned}$$

Тут ми використали факт, що повна варіація згортки двох функцій обмежена добутком їх повних варіацій, дивись теорему 1.3.2(c) в [68]. Ітеруючи цю нерівність робимо висновок, що

$$\begin{aligned} |R_j(t)| &\leq \sum_{k=0}^{j-1} \binom{j}{k} g_{k, j-k}(t) \\ &\leq \sum_{k \leq j/2} \binom{j}{k} \tilde{C}^k g_{0, j-2k}(t) + \sum_{j/2 < k < j} \binom{j}{k} \tilde{C}^{j-k} g_{2k-j, 0}(t), \quad t \in \mathbb{R}. \end{aligned} \quad (2.18)$$

Зауважимо, що $g_{0, j-2k}(t) \leq \mathcal{V}_{\mathbb{R}}(\tilde{\varepsilon}^{*(j-2k)}) \leq (\mathcal{V}_{\mathbb{R}}(\tilde{\varepsilon}))^{j-2k} \leq \tilde{C}_1^{j-2k}$ for $\tilde{C}_1 := \int_{\mathbb{R}} |d\tilde{\varepsilon}(y)| < \infty$. Тому перша сума в правій частині (2.18) задовольняє

$$\begin{aligned} \sum_{k \leq j/2} \binom{j}{k} \tilde{C}^k g_{0, j-2k}(t) &\leq \sum_{k \leq j/2} \binom{j}{k} \tilde{C}^k \tilde{C}_1^{j-2k} \\ &\leq (\tilde{C} \tilde{C}_1^{-1} + \tilde{C}_1)^j \\ &= o \left(\frac{a^j t^j}{j!} \left(1 + \frac{\gamma_0 j}{at} \right)^j \right), \quad t \rightarrow \infty, \end{aligned}$$

оскільки $b^j = b^{j(t)}$ зростає повільніше ніж $\frac{a^j t^j}{j!} \left(1 + \frac{\gamma_0 j}{at}\right)^j$ при $t \rightarrow \infty$, для довільної скінченної константи $b > 0$. Тепер дослідимо другу суму в правій частині (2.18):

$$\begin{aligned} \sum_{j/2 < k < j} \binom{j}{k} \tilde{C}^{j-k} g_{2k-j,0}(t) &= \sum_{j/2 < k < j} \binom{j}{k} \tilde{C}^{j-k} \mathcal{V}_{(-\infty, t]}(\ell^{*(2k-j)}) \\ &= \sum_{j/2 < k < j} \binom{j}{k} \tilde{C}^{j-k} \ell^{*(2k-j)}(t) \\ &= \sum_{j/2 < k < j} \binom{j}{k} \tilde{C}^{j-k} \frac{(at + \gamma_0(2k-j))^{2k-j}}{(2k-j)!} \\ &= \sum_{1 \leq k < j/2} \binom{j}{k} \tilde{C}^k \frac{(at + \gamma_0(j-2k))^{j-2k}}{(j-2k)!}. \end{aligned}$$

Тут друга рівність випливає з монотонності ℓ та третя рівність виконується для достатньо великих t . Для наступних міркувань важливо, що при $k < j/2$ та $t > 0$ виконується

$$\frac{(at + \gamma_0(j-2k))^{j-2k}}{(j-2k)!} \leq \frac{a^{j-2k} t^{j-2k}}{(j-2k)!} \exp\left(\frac{\gamma_0(j-2k)^2}{at}\right).$$

Випадок $\gamma_0 \geq 0$. Отримуємо, при $t > 0$,

$$\begin{aligned} \sum_{1 \leq k < j/2} \binom{j}{k} \tilde{C}^k \frac{(at + \gamma_0(j-2k))^{j-2k}}{(j-2k)!} &\leq \exp\left(\frac{\gamma_0 j^2}{at}\right) \sum_{1 \leq k < j/2} \binom{j}{k} \tilde{C}^k \frac{a^{j-2k} t^{j-2k}}{(j-2k)!} \\ &= \frac{a^j t^j}{j!} \exp\left(\frac{\gamma_0 j^2}{at}\right) \sum_{1 \leq k < j/2} \frac{(j!)^2}{(j-k)!(j-2k)!} \frac{1}{k!} \frac{\tilde{C}^k}{a^{2k} t^{2k}} \\ &\leq \frac{a^j t^j}{j!} \exp\left(\frac{\gamma_0 j^2}{at}\right) \sum_{k \geq 1} j^{3k} \frac{1}{k!} \frac{\tilde{C}^k}{a^{2k} t^{2k}} \\ &= \frac{a^j t^j}{j!} \exp\left(\frac{\gamma_0 j^2}{at}\right) \left(\exp\left(\frac{\tilde{C} j^3}{a^2 t^2}\right) - 1 \right). \end{aligned}$$

Останній множник збігається до нуля, коли $j = j(t) = o(t^{2/3})$, що й потрібно було показати.

Випадок $\gamma_0 < 0$. Міркуючи так само як і у випадку $\gamma_0 \geq 0$, достатньо переви-

рити, що

$$\sum_{1 \leq k < j/2} \frac{1}{k!} \left(\frac{\tilde{C}j^3}{a^2t^2} \right)^k \exp\left(\frac{\gamma_0(j-2k)^2}{at}\right) = o\left(\exp\left(\frac{\gamma_0j^2}{at}\right)\right), \quad t \rightarrow \infty,$$

що рівносильно

$$I_t := \sum_{1 \leq k < j/2} \frac{1}{k!} \left(\frac{\tilde{C}j^3}{a^2t^2} \right)^k \exp\left(\frac{4|\gamma_0|k(j-k)}{at}\right) = o(1), \quad t \rightarrow \infty.$$

Застосовуючи нерівність $\exp\left(\frac{4|\gamma_0|k(j-k)}{at}\right) \leq \exp(4|\gamma_0|a^{-1}k)$, при $1 \leq k < j$, та достатньо великих t , ми робимо висновок, що

$$I_t \leq \sum_{1 \leq k < j/2} \frac{1}{k!} \left(\frac{\tilde{C}j^3 \exp(4|\gamma_0|a^{-1})}{a^2t^2} \right)^k \leq \exp\left(\frac{\tilde{C}j^3}{a^2t^2} \exp(4|\gamma_0|a^{-1})\right) - 1 \rightarrow 0,$$

при $t \rightarrow \infty$.

Доведення завершено. □

Наслідок 32. Нехай $\mathbb{E}\xi^3 < \infty$ та ξ має розподіл абсолютно неперервного типу, тобто розподіл S_k для деякого $k \in \mathbb{N}$ має абсолютно неперервну компоненту. Тоді

$$U^{*(j)}(t) \sim \frac{t^j}{\mu^j j!} \exp\left(\frac{\gamma_0 \mu j^2}{t}\right)$$

при $t \rightarrow \infty$, де $\mu = \mathbb{E}\xi$, $j = j(t)$ – деяка цілочисельна функція, що $j(t) = o(t^{2/3})$ при $t \rightarrow \infty$, та

$$\gamma_0 = \frac{\mathbb{E}\xi^2}{2\mu^2}.$$

Доведення. За накладених на ξ умов з [30, зауваження 3.1.7, с. 121] випливає

$$\int_{[0, \infty]} y \left| d\left(U(y) - \frac{y}{\mu}\right) \right| < \infty.$$

Отже, залишається лише застосувати теорему 31 для функції відновлення.

Константа γ_0 знаходиться у вигляді границі $\lim_{t \rightarrow \infty} (U(t) - \frac{t}{\mu})$, що рівна $\frac{\mathbb{E}\xi^2}{2\mu^2}$, оскільки випадкове блукання неграгчає, див. [2, твердження 6.1, с. 160]. □

Зауважимо, що у спільній роботі [12] містяться більш загальні твердження порівняно з наслідками 30 та 32, а саме розглядаються збурені випадкові блукання й умови на в.в. ξ послаблені.

Зауваження 33. За умов наслідку 32, коли $j(t) = o(t^{1/2})$ при $t \rightarrow \infty$, автоматично виконуються й умови наслідку 30. Проте ніяких протиріч щодо асимптотичної поведінки згортки функцій відновлення в результатах немає, оскільки за таких умов щодо зростання $j(t)$ виконується

$$\exp\left(\frac{\gamma_0 \mu j^2}{t}\right) \sim 1$$

при $t \rightarrow \infty$.

2.3 Висота дерева загального гіллястого процесу

При $t > 0$ визначимо

$$H(t) := \inf\{j \in \mathbb{N} : N_j(t) = 0\}$$

та зауважимо, що $N_j(t) = 0$ м.н. для всіх $j \geq H(t)$. Називатимемо величину $H(t)$ *висотою дерева загального гіллястого процесу* згенерованого збуреним випадковим блуканням T та обмеженого смугою $[0, t]$.

Теорема 34. Для всіх $t \geq 0$ $H(t) < \infty$ м.н. Більш того

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{\gamma} \in (0, \infty) \quad \text{м.н.}, \quad (2.19)$$

де $\gamma := \sup\{z > 0 : \mu(z) < 1\}$ та $\mu(z) := \inf_{s>0} (e^{zs} \frac{\mathbb{E}e^{-s\eta}}{1 - \mathbb{E}e^{-s\xi}})$ при $z > 0$.

Доведення. За припущенням $\mathbb{P}\{\eta = 0\} = 0$. Це означає, що $\lim_{s \rightarrow \infty} \frac{\mathbb{E}e^{-s\eta}}{1 - \mathbb{E}e^{-s\xi}} = 0$ та таким чином

$$\lim_{z \rightarrow 0^+} \mu(z) = 0.$$

Також $\lim_{z \rightarrow \infty} \mu(z) = \lim_{s \rightarrow 0^+} \frac{\mathbb{E}e^{-s\eta}}{1 - \mathbb{E}e^{-s\xi}} = \infty$. Це показує, що $\gamma \in (0, \infty)$.

Нагадаємо, що, при $n \in \mathbb{N}$, $(T_r^{(n)})_{r \in \mathbb{N}}$ позначає деяку нумерацію часів народження в n -у поколінні загального гіллястого процесу. Покладемо

$B(n) := \inf_{r \geq 1} T_r^{(n)}$. Згідно з результатом Бігінса (наслідок на с. 635 в [8]),

$$\lim_{n \rightarrow \infty} \frac{B(n)}{n} = \gamma \quad \text{м.н.}, \quad (2.20)$$

де γ визначається аналогічно до умови теореми, проте $\mu(z) = \inf_{s > 0} (e^{zs} m(s))$, де

$$m(s) := \mathbb{E} \left[\sum_{k \geq 1} e^{-sT_k} \right].$$

Для збуреного випадкового блукання T легко підрахувати, що $m(s) = \frac{\mathbb{E}e^{-s\eta}}{1 - \mathbb{E}e^{-s\xi}}$.

Оскільки, для $n \in \mathbb{N}$ та $t > 0$, $\{H(t) > n\} = \{B(n) \leq t\}$ й, відповідно до (2.20), $\lim_{n \rightarrow \infty} B(n) = +\infty$ м.н., ми робимо висновок, що $H(t) < \infty$ м.н.

Нарешті, маємо $B(H(t)) > t \geq B(H(t) - 1)$ м.н. Ліва частина нерівності забезпечує $\lim_{t \rightarrow \infty} H(t) = +\infty$ м.н., що разом з (2.20) доводить (2.19) за теоремою про двох поліцейських. \square

Досить рідко можна явно підрахувати константу γ . Ось один щасливий виняток. Нехай $(\xi, \eta) = (-\log u, -\log(1 - u))$, де u має рівномірний розподіл на $[0, 1]$. Тоді розподіл послідовності $(e^{-T_i})_{i \in \mathbb{N}}$ відомий як *розподіл Гріффітса-Енгена-Макклоскі* з параметром 1.

В цьому випадку маємо

$$\mathbb{E}e^{-s\eta} = \mathbb{E}e^{-s\xi} = \frac{1}{s + 1}$$

та $m(s) = \frac{1}{s}$. Тоді мінімум $e^{zs} \frac{1}{s}$ можна підрахувати за допомогою похідної й він досягається при $s = \frac{1}{z}$. Остаточно $\mu(z) = ez$ при $z > 0$, що дає $\gamma = e^{-1}$.

2.4 Математичне моделювання

Спочатку перевіримо збіжність висоти $H(t)$ дерева загального гіллястого процесу побудованого за збуреним випадковим блуканням. Теорема 34 стверджує збіжність майже напевне, що означає, що з ймовірністю 1 послідовність має збігатись для кожної реалізації гіллястого процесу.

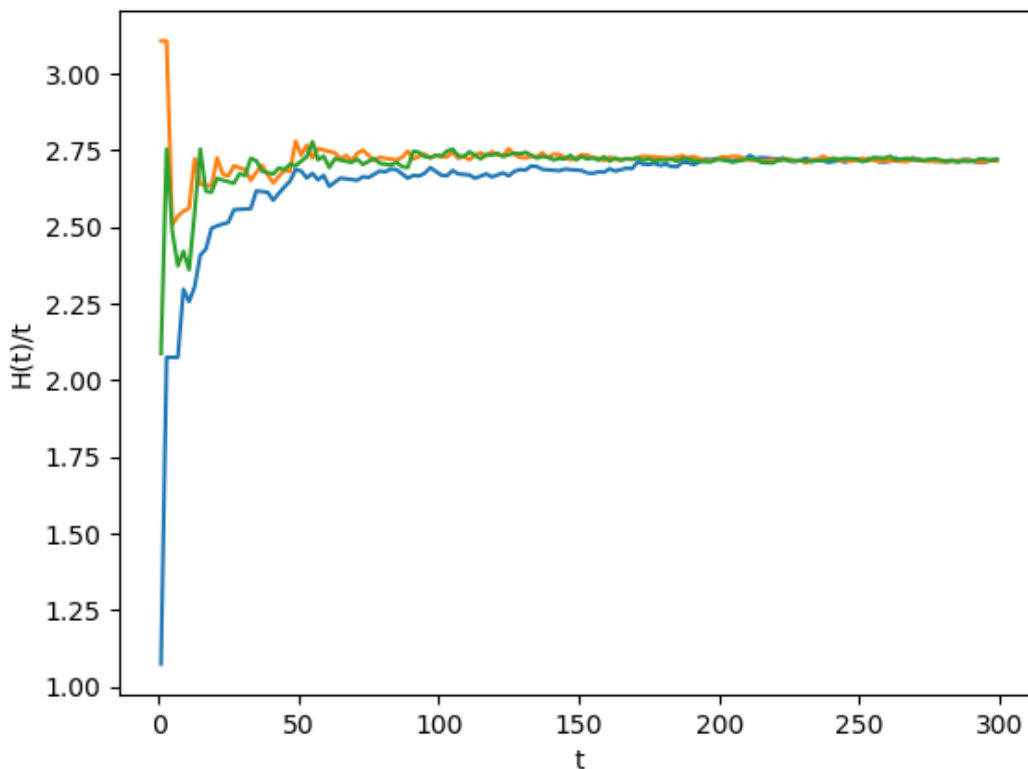


Рис. 2.2: Перший приклад еволюції висоти дерева загального гіллястого процесу за трьох реалізацій процесу.

Згідно до зауваження після теореми, у випадку, коли $(\xi, \eta) = (-\log u, -\log(1 - u))$, а u має рівномірний на $[0, 1]$ розподіл, константа γ була підрахована, а саме має виконуватись

$$\frac{H(t)}{t} \rightarrow e \quad \text{м.н.}$$

при $t \rightarrow \infty$. На рисунку 2.2 одночасно зображено три реалізації гіллястого процесу та відповідну еволюцію відношення $H(t)/t$ з плином часу.

Без підрахунку явного вигляду величини γ цікаво побачити чи виконується збіжність до певної константи для інших випадкових величин. Так на рисунку 2.3 зображена еволюція висоти дерева трьох реалізацій процесу, коли ξ має хі-квадрат розподіл з одним ступенем свободи, а η має рівномірний на $[0, 1]$ розподіл й не залежить від ξ .

Тепер розглянемо згортки функцій відновлення, а саме твердження на-

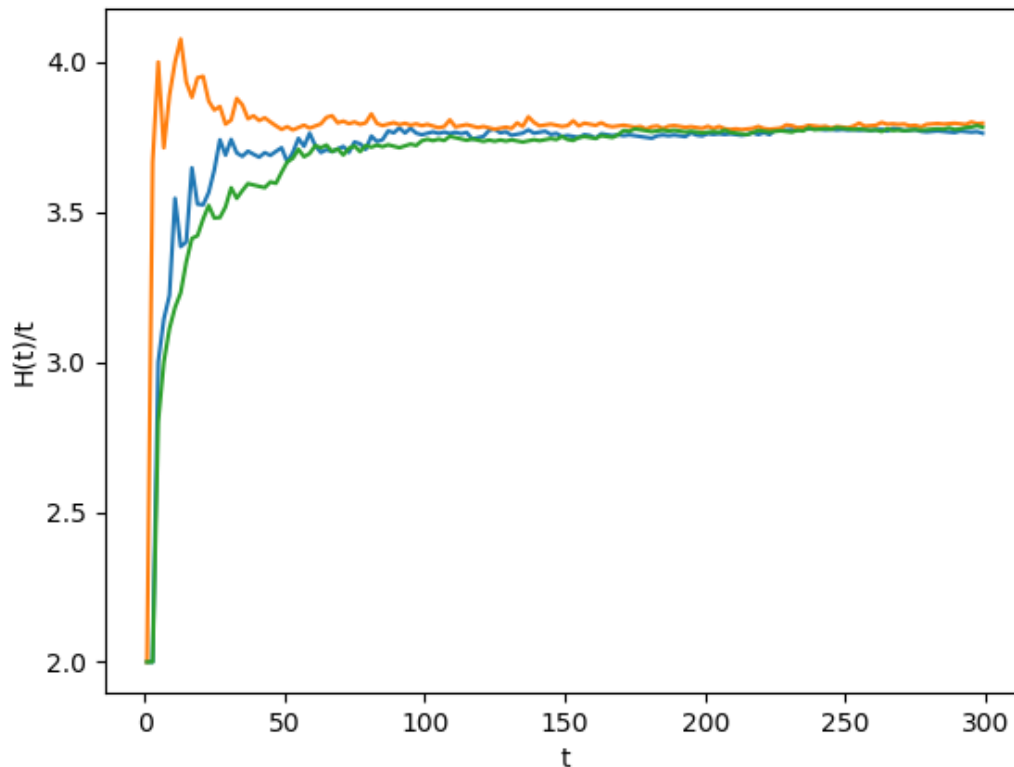


Рис. 2.3: Другий приклад еволюції висоти дерева загального гіллястого процесу за трьох реалізацій процесу.

слідків 30 та 32. Досить легко вибрати розподіл випадкової величини ξ так, щоб задовольнялись умови обох тверджень, проте другий наслідок стверджує різку зміну поведінки згортки, коли $j(t)$ стає зростати швидше, ніж $t^{1/2}$.

Нехай ξ має розподіл Ерланга 2-го порядку з параметром λ , тоді функція відновлення матиме вигляд (див. [4])

$$U(t) = \frac{2\lambda t + 3 + \exp(-2\lambda t)}{4}, \quad t \geq 0.$$

Покладемо $\lambda = 1$ та розглянемо кратність згортки у вигляді $j(t) = \lfloor t^\alpha \rfloor$, де параметр $\alpha \in (0, 1)$ буде змінюватись. На рисунку 2.4 зображено еволюцію нормованої згортки функції відновлення при зростанні часу t , а на осі абсцис розглянуті різні значення параметра α . Під поняттям нормованої згортки розуміється наступне відношення:

$$U^{*(j)}(t) \frac{\mu^j j!}{t^j}.$$

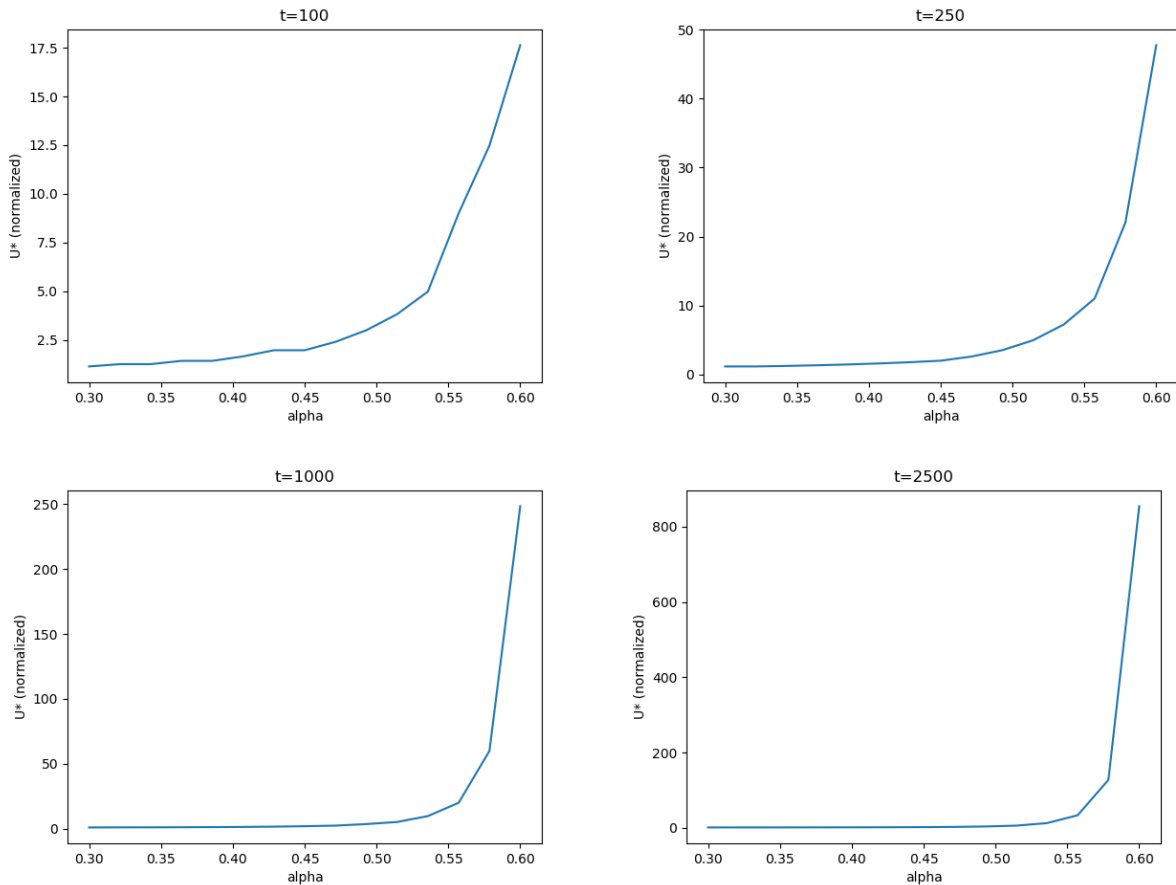


Рис. 2.4: Еволюція згортки функцій відновлення при зростанні часу t за різних значень α .

Таким чином це відношення має збігатись до 1 при $\alpha < 1/2$, та вести себе як експонента інакше. Графіки це підтверджують, зокрема можна бачити різку розбіжність при збільшенні t . За рахунок цієї розбіжності важко побачити збіжність до 1 при $\alpha < 1/2$, тому на рисунку 2.5 окремо показана область збіжності.

Зауваження 35. Розподіл випадкової величини ξ було обрано таким чином, щоб функція відновлення не була тривіальною, як у випадку експоненційного розподілу, проте була вже явно підрахованою аналітично. Останнє використано з метою усунення додаткових похибок при моделюванні й підрахунку самих згортки, оскільки похибка в підрахунку функції відновлення буде відсутня. Сам розподіл Ерланга $(m - 1)$ -го порядку виникає при дослідженні m -арних дерев пошуку за допомогою гіллястих процесів ([34]).

Зауваження 36. Явний або близький до явного вигляд функції відновлення

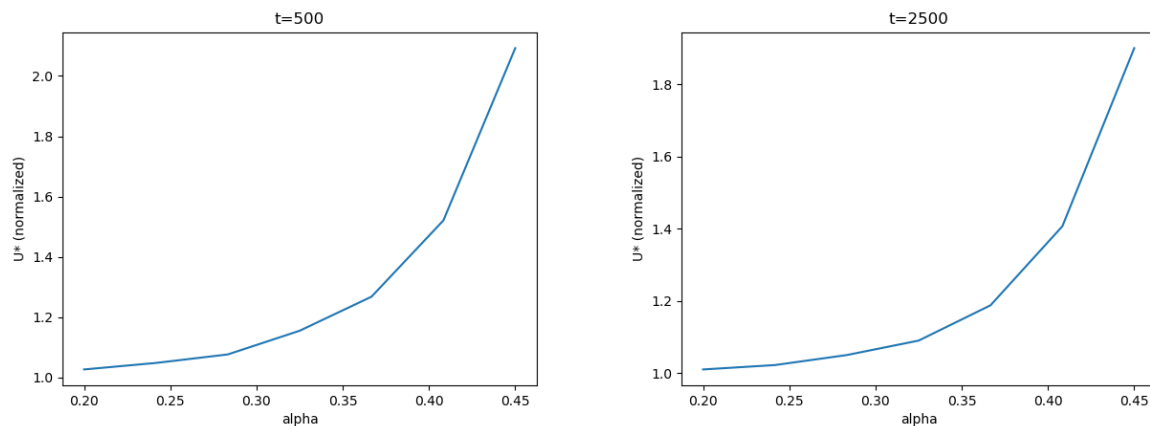


Рис. 2.5: Еволюція згорток функцій відновлення при зростанні часу t , де параметр α менше $1/2$.

підраховано лише для невеликої кількості розподілів ξ . У випадках, коли ці значення потрібно також знаходити, ми посилаємось до методів, описаних в роботах [53, 69, 78]. Всі ці методи мають різну швидкодію та знаходять значення з різною похибкою, що в свою чергу можуть різнитись в залежності від самого розподілу ξ .

2.4.1 Програмний код мовою Python3. Спочатку наведемо код для моделювання дерева загального гіллястого процесу та підрахунку його висоти. В лістингу 2.1 наведено дві основні функції, що будуть необхідні для основної частини коду. Функція `get_xi_eta()` повертає пару значень (ξ, η) , а функція `generate_T()` відповідно генерує реалізацію точкового процесу T до деякого моменту.

```

1 # height.py
2 import random
3 import math
4 import bisect
5 import matplotlib.pyplot as plt
6
7
8 def get_xi_eta():
9     # get pair of main random variables
10    w = random.normalvariate(0, 1)
11    xi = w * w
12    eta = random.uniform(0, 1)
13    return xi, eta

```

```

14
15
16 def generate_T(shift=0, max_t=10):
17     # generate point process T up to position max_t shifted by value shift
18     S_k_1 = 0
19     T_k = 0
20     T = []
21     while T_k < max_t:
22         xi, eta = get_xi_eta()
23         T_k = S_k_1 + eta
24         T.append(T_k + shift)
25         S_k_1 += xi
26     return T

```

Лістинг 2.1: Основні функції для генерування дерева загального гіллястого процесу.

Одразу зауважимо, що генерувати кожне покоління гіллястого процесу просто неможливо через те, що кількість осіб в кожному поколінні зростає експоненційно, до того ж ще й треба обирати досить великий момент часу t для котрого проводити дослідження висоти, щоб спостерігати збіжність. Тому було прийнято рішення обмежити кількість осіб, що народжуються від кожного нащадка, та кількість осіб, що дають подальше потомство в кожному поколінні. На кожній ітерації при цьому залишаються в живих лише певна кількість осіб з найменшим часом народження, бо саме ці особи впливають на підрахунок висоти дерева. В лістингу 2.2 наведено основну частину для моделювання й підрахунку висоти дерева загального гіллястого процесу.

```

1 # height.py
2
3 # ...
4
5 T_LENGTH = 25 # max length of each point process T we generate
6 GENERATION_MINS_ALIVE = 500 # how many min individuals to keep alive
7                               # in each generation
8 MAX_TIME = 300 # up to which time we compute height
9 PROCESSES = 3 # how many different process realizations we generate
10
11
12 for _ in range(PROCESSES):
13     generations = [[0]]
14
15     while generations[-1][0] < MAX_TIME:

```

```

16     current_generation = []
17     for ancestor in generations[-1]:
18         T = generate_T(ancestor, T_LENGTH)
19         current_generation += T
20     current_generation.sort()
21     generations.append(current_generation[:GENERATION_MINS_ALIVE])
22
23     B_n = []
24     for generation in generations:
25         B_n.append(min(generation))
26
27     x = []
28     y = []
29     for t in range(1, MAX_TIME, 2):
30         h = bisect.bisect_right(B_n, t) # binary search
31         x.append(t)
32         y.append(h / t)
33
34     plt.plot(x, y)
35
36     plt.xlabel("t")
37     plt.ylabel("H(t)/t")

```

Лістинг 2.2: Основний код для генерування дерева загального гіллястого процесу й підрахунку його висоти.

В цьому лістингу константи `T_LENGTH` та `GENERATION_MINS_ALIVE` обмежують два вищенаведених параметри відповідно. Зауважимо, що також в процесі моделювання була перевірена поведінка висоти за збільшення цих параметрів, проте це не вплинуло на характер збіжності, що підтверджує стійкість подібної евристики для підрахунку висоти дерева з практичної точки зору. Додатково зауважимо, що для підрахунку висоти достатньо знати лише мінімальні часи народження осіб в кожному поколінні, при чому ця послідовність не спадає, тому значення висоти можна ефективно знаходити за допомогою бінарного пошуку.

В лістингу 2.3 наведено програмний код для підрахунку згорток функцій відновлення та побудови графіків, що наведені на рисунках 2.4 та 2.5. Інтегрування виконано методом Сімпсона, також звернемо увагу, що окремо враховується доданок, що виникає за рахунок того, що функція відновлення

має стрибок в нулі. Значення згорток окремо підраховуються в усіх точках рівномірної ґратки, оскільки всі ці значення знадобляться на наступній ітерації. Нагадаємо, що підраховуються саме значення нормованої згортки, де нормування доповнюється на кожній ітерації. Це необхідно для того, щоб значення згорток залишались малими, що важливо для точності підрахунків при використанні чисел з рухомою комою.

```

1 # conv.py
2 import numpy as np
3 import scipy.integrate
4 import matplotlib.pyplot as plt
5
6
7 def U(t, _lambda=1):
8     # renewal function with jumps distributed as Erlang with 2 stages
9     return (t >= 0) * (2 * _lambda * t + 3 + np.exp(-2 * _lambda * t)) / 4
10
11
12 MAX_T = 1000 # max value of t
13 INTERVALS = int(MAX_T) * 10 # intervals for integration
14 mu = 2 # expectation of xi
15
16 ALPHA_MIN = 0.3 # min value of power alpha
17 ALPHA_MAX = 0.6 # max value of power alpha
18 ALPHA_INTERVALS = 15 # how many different alphas to compute
19 MAX_J = int(MAX_T ** ALPHA_MAX) # max value of j (to stop computations)
20
21 t = np.linspace(0, MAX_T, INTERVALS + 1, dtype="float64")
22
23 U_1 = U(t) # renewal function computed in fixed points
24
25 U_convolution = [None] # list of convolutions, 0th value is ignored
26 U_convolution.append(U_1 / MAX_T * mu) # adding normalized 1st convolution
27
28 for c in range(2, MAX_J + 1):
29     conv = np.zeros(INTERVALS + 1, dtype="float64")
30     for i in range(INTERVALS + 1):
31         U_prev = U_convolution[-1][i+1]
32         # previous convolution truncated up to current value t_i
33         res = scipy.integrate.simps(y=U_prev[:, -1], x=U_1[i+1])
34         # simpson integration of int y dx, U_prev is flipped
35
36         res += U_1[0] * U_prev[-1]

```

```

37         # adding a missed value due to jump of renewal function at t=0
38
39         conv[i] = res
40         conv /= MAX_T / mu / c # normalization
41         U_convolution.append(conv)
42
43 alpha = np.linspace(ALPHA_MIN, ALPHA_MAX, ALPHA_INTERVALS)
44 U_j = np.zeros(ALPHA_INTERVALS, dtype="float64")
45 for i, a in enumerate(alpha):
46     j = int(MAX_T ** a)
47     U_j[i] = U_convolution[j][-1]
48
49 plt.xlabel("alpha")
50 plt.ylabel("U* (normalized)")
51 plt.title("t=%s" % MAX_T)
52 plt.plot(alpha, U_j)

```

Лістинг 2.3: Програмний код для підрахунку згорток функцій розподілу.

2.5 Висновки до розділу 2

В цьому розділі дисертаційної роботи було зроблено внесок в теорію дерев загальних гіллястих процесів. Зокрема, були отримані наступні результати:

- Отримана асимптотична поведінка згорток функцій лінійного росту у випадку, коли кількість згорток також зростає.
- Загальні результати для цих функцій використані для виведення асимптотичної поведінки згорток функцій відновлення.
- Досліджена асимптотична поведінка висоти дерева загального гіллястого процесу, побудованого на збуреному випадковому блуканні.
- Всі результати були підтвержені за допомогою математичного моделювання.

Зауважимо, що в спільній роботі [12] наводиться набагато більше результатів, отриманих в цьому напрямку, що належать іншим авторам.

Серед напрямків подальших досліджень можна виділити наступні:

- Теорія дерев загальних гіллястих процесів все ще не є повною. Для характеристик збурених гіллястих блукань залишаються відкритими питання щодо сильніших типів збіжності, наприклад, збіжність кількості живих осіб в j -му поколінні майже напевно, коли $j = j(t)$ зростає з часом. Іншим напрямком розвинення теорії є розгляд інших типів випадкових блукань.
- Залишаються відкритими питання щодо асимптотичної поведінки кратних згорток Лебега-Стілтєса функцій лінійного росту, коли кількість елементів згорток $j = j(t)$ зростає швидше ніж $t^{2/3}$.
- Вже отримані результати можна застосовувати для дослідження властивостей більш широкого класу дерев, наприклад, для m -арних дерев пошуку, а не тільки випадкових рекурсивних дерев.

Розділ 3

VP-дерева

3.1 Означення та обговорення

Нехай $(\mathcal{X}, \text{dist})$ – довільний метричний простір. VP-деревом називають бінарне дерево, в якому в кожній вершині T зберігається два значення: елемент $x_T \in \mathcal{X}$, що називають опорною точкою, та порогове значення $\tau_T > 0$. Одним з найбільш загальних способів побудувати VP-дерево є рекурсивний підхід за допомогою послідовності елементів x_1, x_2, \dots . Дерево порожнє на початку. Перший елемент x_1 завжди зберігається в його корені тазначається деяке порогове значення τ . Для того, щоб вставити n -ий елемент x_n виконується наступна рекурсивна процедура. Ми починаємо з вершини T , що є коренем дерева. Якщо $\text{dist}(x_T, x_n) \leq \tau$, то ми продовжуємо рекурсивно, додаючи новий елемент до лівого сина T , інакше ми рухаємось в праве піддерево. Коли на деякому кроці вершина порожня, тоді величина x_n зберігається в ній та їй присвоюється деяке порогове значення.

В цьому розділі ми розглядаємо окремий клас випадкових VP-дерев. Більш точно, ми розглядаємо метричний простір $([-1, 1]^d, \text{dist})$ ($d \geq 1$) з функцією відстані

$$\text{dist}(x, y) := \ell_\infty(x, y) = \max_{1 \leq i \leq d} |x_i - y_i|.$$

Ми також повинні вказати вибір порогових значень. До кінця цього розділу дисертаційної роботи зафіксуємо параметр $\tau \in (0, 1)$. Порогове значення

кожної вершини T вибирається за правилом $\tau_T = \tau^h$, де h позначає відстань від T до кореня дерева плюс 1. Іншим можливим способом буде обирати h як кількість “поворотів вліво” в шляху від кореня до T плюс 1. Зауважимо, що такий вибір порогових значень досить природній. Якщо хтось захоче обирати їх як медіани відстаней від опорної точки, то порогові значення також будуть слідувати експоненційному шаблону.

Визначимо *найлівіший шлях* VP-дерева (чи довільного бінарного дерева) як шлях, що починається в корені та завжди йде до лівого сина в поточній вершині поки не дійде до листка. Зауважимо, що для обох вищеназваних правил вибору порогового значення, величина h однакова для елементів найлівішого шляху.

Послідовність елементів x_1, x_2, \dots будемо вважати послідовністю незалежних рівномірно розподілених точок в $[-1, 1]^d$, що робить VP-дерево, яке ми будуємо, випадковим.

Кожна вершина в найлівішому шляху може бути пронумерована, починаючи з 1. Щоб зрозуміти як формується та еволюціонує найлівіший шлях пов’яжемо з кожною вершиною T_h ($h \geq 1$) в цьому шляху частину простору $I_h \subset \mathcal{X}$, що містить всі точки x такі, що при додаванні в дерево були б збережені в лівому піддереві T_h . Очевидно, що $I_h \subset I_{h-1}$ та, більш того, точка x буде збережена в лівому піддереві T_h , якщо $\text{dist}(x, x_{T_h}) \leq \tau^h$. Це дає

$$I_h = I_{h-1} \cap B_{\tau^h}(x_{T_h}), \quad h \geq 1, \quad (3.1)$$

де $B_{\tau^h}(x_{T_h})$ позначає метричну кулю радіусу τ^h та центром в точці x_{T_h} , тобто

$$B_{\tau^h}(x_{T_h}) = \{x \in \mathcal{X} : \text{dist}(x, x_{T_h}) \leq \tau^h\}.$$

Зауважимо, що за нашим вибором метрики, $\text{dist}(x, x_{T_h}) \leq \tau^h$ представляє d -вимірний куб. Корінь T_1 дерева завжди є частиною найлівішого шляху, тому можемо природно покласти $I_0 := \mathcal{X} = [-1, 1]^d$, що теж є кубом. З (3.1) робимо висновок, що, для всіх $h \geq 1$, I_h є перетином прямокутного паралелепіпеда та куба, а тому й саме є прямокутним паралелепіпедом зі сторонами

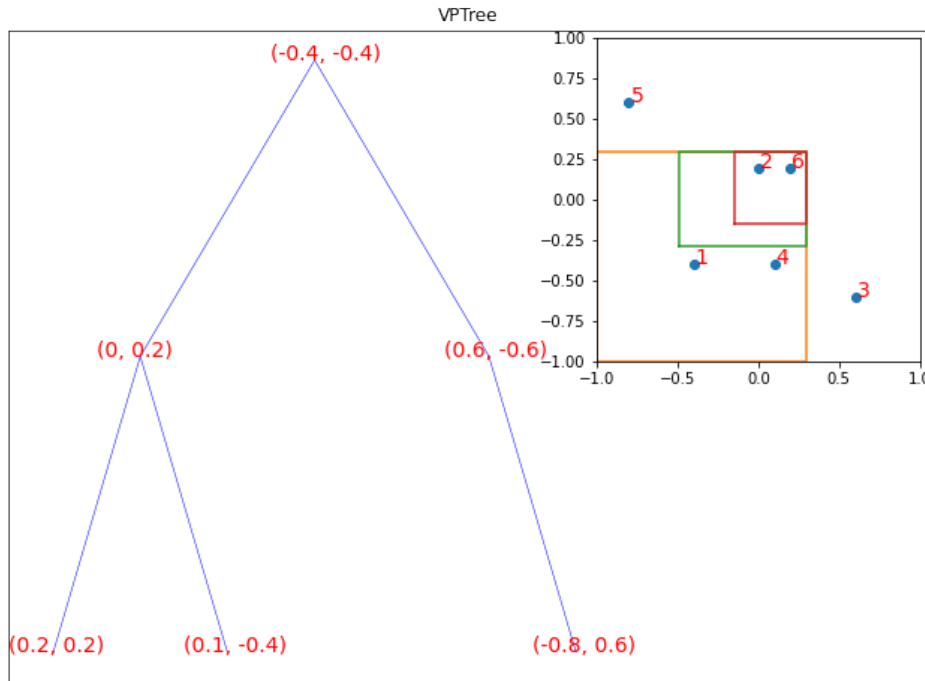


Рис. 3.1: Зображення VP-дерева, отриманого після додавання 6 точок з $\tau = 0.7$. У правому верхньому куті наведені розбиття простору, що утворюють перша, друга та шоста точки як точки найлівішого шляху.

паралельними осям координат. Для таких фігур достатньо (з точністю до паралельних переносів) зберігати лише інформацію про їх розміри по кожній розмірності, які позначимо через $I_h^{(i)}$, $1 \leq i \leq d$.

Для кращого представлення на рисунку 3.1 зображено приклад реалізації дерева, побудованого на шести послідовних точках, самі точки та розбиття простору, що породжують точки найлівішого шляху.

Нехай T_{h-1} – поточна остання вершина в найлівішому шляху. Тоді нова вершина T_h буде додана зліва від T_{h-1} , якщо нова точка x_n потрапить всередину I_{h-1} . З властивостей рівномірного розподілу за умови, що $x_n \in I_{h-1}$, x_n має рівномірний розподіл в I_{h-1} . Беручи до уваги, що I_{h-1} прямокутний паралелограм, також отримуємо, що розташування x_n всередині I_{h-1} має рівномірний розподіл по кожній розмірності $i = 1, \dots, d$. Отже, можна записати

d рекурсивних рівнянь по кожній розмірності для розмірів кожної сторони I_h , що є просто розміром перетину двох відрізків: відрізка довжини $I_{h-1}^{(i)}$ та відрізка радіусу τ^h і з центром розподіленим рівномірно всередині першого відрізка. Ми підраховуємо розмір $I_h^{(i)}$ як відстань між правим та лівим кінцями перетину. Нехай $u_h^{(i)}$, $h \geq 1$, $1 \leq i \leq d$, – сім'я незалежних рівномірно розподілених випадкових величин на $[0, 1]$. Маємо, для $h \geq 1$, $1 \leq i \leq d$,

$$\begin{aligned}
I_h^{(i)} &= \min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - \max\{0, u_h^{(i)} I_{h-1}^{(i)} - \tau^h\} \\
&= (u_h^{(i)} I_{h-1}^{(i)} - \max\{0, u_h^{(i)} I_{h-1}^{(i)} - \tau^h\}) \\
&\quad + (\min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - u_h^{(i)} I_{h-1}^{(i)}) \\
&= (u_h^{(i)} I_{h-1}^{(i)} + \min\{0, \tau^h - u_h^{(i)} I_{h-1}^{(i)}\}) \\
&\quad + (\min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - u_h^{(i)} I_{h-1}^{(i)}) \\
&= \min\{u_h^{(i)} I_{h-1}^{(i)}, \tau^h\} + \min\{(1 - u_h^{(i)}) I_{h-1}^{(i)}, \tau^h\}.
\end{aligned} \tag{3.2}$$

Зробивши заміну $X_h^{(i)} := I_h^{(i)} / \tau^h$, отримуємо основну рекурсивну формулу, що описує еволюцію I_h :

$$X_h^{(i)} = \min\left\{\frac{u_h^{(i)} X_{h-1}^{(i)}}{\tau}, 1\right\} + \min\left\{\frac{(1 - u_h^{(i)}) X_{h-1}^{(i)}}{\tau}, 1\right\}. \tag{3.3}$$

3.2 Збіжність довжини найлівішого шляху

3.2.1 Аналіз процесу (X_h) . Процес (3.3) визначає розміри $X_h = I_h / \tau^h$ в різних розмірностях. $(X_h)_{h \geq 0}$ може розглядатись як ланцюг Маркова на паралелепіпедах (опускаючи їх розташування). Ми позначаємо через $|I_h|$ (чи $|X_h|$) міру Лебега відповідного паралелепіпеда, котра може бути легко підрахована знаючи розміри кожного ребра

$$|I_h| = \prod_{i=1}^d I_h^{(i)}. \tag{3.4}$$

Твердження 37. Ланцюг Маркова (3.3) задовольняє

$$1 \leq X_h^{(i)} \leq 2,$$

при $h \geq 0$ та $1 \leq i \leq d$.

Доведення. Зафіксуємо i та доведемо твердження по h , використовуючи математичну індукцію. Для випадкового VP-дерева $I_0 = \mathcal{X} = B_1(0) = [-1, 1]^d$, тому $X_0^{(i)} = I_0^{(i)} = 2$ й твердження справджується. Оскільки $\min\{a, b\} \leq b$, то з (3.3) маємо

$$X_h^{(i)} \leq 1 + 1 = 2.$$

Щоб показати, що $X_h^{(i)} \geq 1$ спочатку зауважимо, що якщо будь-який з мінімумів в (3.3) приймає значення 1, то твердження дійсне. З іншого боку, якщо обидва мінімуми менші за 1, то

$$X_h^{(i)} = \frac{u_h^{(i)} X_{h-1}^{(i)}}{\tau} + \frac{(1 - u_h^{(i)}) X_{h-1}^{(i)}}{\tau} = \frac{X_{h-1}^{(i)}}{\tau} \geq X_{h-1}^{(i)} \geq 1,$$

за припущенням індукції та оскільки $\tau \in (0, 1)$. □

Позначимо через S множину усіх можливих значень, що ланцюг (X_h) може прийняти за умови, що $X_0 = B_1(0)$. Зауважимо, що S незліченна.

Ми будемо використовувати поняття стохастичного порядку, означення якого наводимо нижче.

Означення 38. Для двох випадкових величин X та Y кажемо, що Y стохастично домінує X та позначаємо $X \stackrel{st}{\leq} Y$, якщо

$$\mathbb{P}\{X \leq t\} \geq \mathbb{P}\{Y \leq t\} \quad \text{для всіх } t \in \mathbb{R}.$$

Властивості стохастичного порядку, зокрема ті, що будуть використовуватись в подальшому в цьому розділі, можна знайти в книзі [72].

Ключовим інгредієнтом подальшого аналізу є наступна теорема, що стверджує, що ланцюг $(X_h)_{h \geq 0}$ відвідує стан $B_1(0)$ відносно часто.

Теорема 39. *Зафіксуємо $\alpha \in S$ та покладемо*

$$R_\alpha := \inf\{h \geq 1 : X_h = B_1(0) \text{ за умови, що } X_0 = \alpha\}.$$

Тоді $\mathbb{E}R_\alpha < \infty$.

Зауваження 40. Ми вважаємо, що аналог цієї теореми виконується для довільної метрики в $[-1, 1]^d$, проте нам не вдалось довести її в такій загальності.

Доведення теореми 39. Перепишемо (3.1) наступним чином:

$$X_{h+1} = \tau^{-1} X_h \cap B_1(u_{h+1}^*), \quad (3.5)$$

де u_{h+1}^* рівномірно розподілена в.в. всередині X_h/τ .

Оскільки розташування паралелепіпедів неважливе, рекурсивну рівність (3.5) можна ще переписати як

$$X_{h+1} = \frac{X_h - u_{h+1}}{\tau} \cap B_1(0), \quad (3.6)$$

де u_{h+1} рівномірно розподілена в.в. в X_h .

З твердження 37 випливає, що для будь-якого стану X_h існує d -вимірний куб зі сторонами 1 всередині X_h , чи, іншими словами, завжди існує точка \hat{x}_h така, що $B_\delta(\hat{x}_h) \subset X_h$ з $\delta = 1/2$.

Припустимо, що точка u_{h+1} була обрана всередині $B_\delta(\hat{x}_h)$ та позначимо

$$\Delta = \text{dist}(u_{h+1}, \hat{x}_h) \leq \delta.$$

За нерівністю трикутника отримуємо

$$\max_{y \in \partial B_\delta(\hat{x}_h)} \text{dist}(u_{h+1}, y) \leq \Delta + \delta,$$

$$\min_{y \in \partial B_\delta(\hat{x}_h)} \text{dist}(u_{h+1}, y) \geq \delta - \Delta,$$

де ∂B позначає границю множини B .

Тоді, з огляду на (3.6), маємо

$$\max_{y \in \partial((B_\delta(\hat{x}_h) - u_{h+1})/\tau)} \text{dist}(0, y) \leq \frac{\Delta + \delta}{\tau}, \quad (3.7)$$

та

$$\min_{y \in \partial((B_\delta(\hat{x}_h) - u_{h+1})/\tau)} \text{dist}(0, y) \geq \frac{\delta - \Delta}{\tau}. \quad (3.8)$$

З (3.7) отримуємо першу властивість: якщо $\frac{\Delta + \delta}{\tau} \leq 1$ (чи еквівалентно $\Delta \leq \tau - \delta$), тоді $B_{\delta/\tau}((\hat{x}_h - u_{h+1})/\tau) \subset X_{h+1}$. З (3.8) отримуємо другу властивість: якщо $\frac{\delta - \Delta}{\tau} \geq 1$ (чи рівносильно $\Delta \leq \delta - \tau$), то $X_{h+1} = B_1(0)$.

Зафіксуємо константу $k := \min\{k \in \mathbb{N} : \frac{1+\tau}{2}\tau^{k-1} \leq \frac{1}{2}\}$.

Оскільки $B_{1/2}(\hat{x}_h) \subset X_h$, то маємо $B_{\frac{1+\tau}{2}\tau^{k-1}}(\hat{x}_h) \subset X_h$. Оскільки $k \geq 2$, то виконується $\frac{1+\tau}{2}\tau^{k-1} < \tau$. Таким чином, якщо $u_{h+1} \in B_{\varepsilon_1}(\hat{x}_h)$ з $\varepsilon_1 = \tau - \frac{1+\tau}{2}\tau^{k-1}$, що відбувається з додатною ймовірністю, отримаємо $B_{\frac{1+\tau}{2}\tau^{k-2}}(\hat{x}_{h+1}) \subset X_{h+1}$ за першою властивістю. Повторюючи цей процес $k-1$ разів сумарно, де $\varepsilon_i = \tau - \frac{1+\tau}{2}\tau^{k-i} > 0$, при $1 \leq i \leq k-1$, отримуємо, що $B_{\frac{1+\tau}{2}}(\hat{x}_{h+k-1}) \subset X_{h+k-1}$ та $\frac{1+\tau}{2} > \tau$. Отже, якщо $u_{h+k} \in B_{\varepsilon_k}(\hat{x}_{h+k-1})$, що аналогічно відбувається з додатною ймовірністю, де $\varepsilon_k = \frac{1+\tau}{2} - \tau$, отримуємо, що $X_{h+k} = B_1(0)$ за другою властивістю.

Підсумовуючи, бачимо, що для довільного стану $\alpha \in S$ ланцюг перестрибує за k кроків в стан $B_1(0)$ з додатною ймовірністю, що не залежить від початкового стану α . Зауважимо, що k залежить лише від τ , тобто є наперед фіксованою константою.

Розіб'ємо еволюцію ланцюгу в блоки довжини k . Стан на кінці кожного блоку може бути $B_1(0)$ з додатною ймовірністю, позначимо її як p , що отримана з вищеописаної схеми. Позначимо через g номер першого блоку, що завершується в стані $B_1(0)$. Тоді оскільки реальна ймовірність бути в стані $B_1(0)$ наприкінці блоку не менша за p , то маємо, для всіх $t \geq 0$,

$$\mathbb{P}\{g \leq t\} \geq \mathbb{P}\{G \leq t\},$$

де G має геометричний розподіл з параметром p (кількість випробувань). Чи в термінах стохастичного порядку

$$g \stackrel{st}{\leq} G.$$

Аналогічно, оскільки момент R_α першого потрапляння до стану $B_1(0)$ (після початкового стану) може відбутись і всередині кожного блоку, то ми можемо записати

$$R_\alpha \stackrel{st}{\leq} kg.$$

З властивостей стохастичного порядку, нерівність також зберігається для математичних сподівань, тому, поєднуючи останні два порівняння, отримує-

МО

$$\mathbb{E}R_\alpha \leq k\mathbb{E}G = \frac{k}{p} < \infty.$$

□

Зауваження 41. Звернемо увагу на те, що в доведенні теореми 39 ми не використовували істотно, що u_h має рівномірний розподіл, та доведення зберігається для будь-якого абсолютно неперервного розподілу, оскільки ймовірність, що u_h потрапить всередину деякої кулі $B_\varepsilon(\hat{x}_h)$ буде додатною та більшою за деяку фіксовану константу.

Зауваження 42. Процес (3.5) в окремому випадку $\tau = 1$ носить назву спадного процесу Балінта Тота. Цей процес був досліджений в різноманітних окремих випадках, зокрема, в розмірності один в роботі [1], а у вищих розмірностях з $B_1(0) = [-1, 1]^d$ в [45], та в інших метриках в [5, 45]. Проте еволюція процесу з $\tau < 1$ сильно відрізняється й потребує незалежного вивчення, до того ж методи, що використовувались для $\tau = 1$ не можуть бути застосовані у випадку $\tau < 1$.

З теореми 39 випливає, що ланцюг $(X_h)_{h \geq 0}$ еволюціонує наступним чином. Існує спеціальний стан $B_1(0)$, котрий ланцюг відвідує нескінченно часто й час повернення до цього стану м.н. скінченний й, більш того, математичне сподівання часу повернення скінченне. Отже, вся еволюція X_h може бути розбита на незалежні цикли між відвідуваннями стану $B_1(0)$.

Для продовження нам знадобиться поняття ланцюга Гарріса. Для додаткової інформації щодо ланцюгів Гарріса та прикладів ми посилаємось до книги [28, глава 6]. Наведемо тут лише деякі необхідні означення.

Означення 43. Ланцюг маркова $(X_h)_{h \geq 0}$ з простором станів S називають ланцюгом Гарріса, якщо існують дві множини $A, B \subset S$, додатна функція $q(x, y) \geq \varepsilon > 0$ при $x \in A, y \in B$, та ймовірнісна міра ρ сконцентрована на B такі, що наступні дві умови виконуються:

- (1) $\mathbb{P}\{\inf\{h \geq 0 : X_h \in A\} < \infty\} > 0$ для всіх можливих початкових станів $X_0 \in S$;

(2) якщо $x \in A$ та $C \subset B$, то $\mathbb{P}\{X_{h+1} \in C | X_h = x\} \geq \int_C q(x, y) \rho(dy)$.

Означення 44. Ланцюг Гарріса називається рекурентним, якщо існує стан α такий, що $\mathbb{P}\{\inf\{h \geq 1 : X_h = \alpha\} < \infty\} = 1$, коли $X_0 = \alpha$.

Означення 45. Рекурентний ланцюг Гарріса називається аперіодичним, якщо найбільший спільний дільник множини $\{h \geq 1 : \mathbb{P}\{X_h = \alpha | X_0 = \alpha\} > 0\}$ рівний 1.

Твердження 46. $(X_h)_{h \geq 0}$ є аперіодичним рекурентним ланцюгом Гарріса.

Доведення. Спочатку покажемо, що $(X_h)_{h \geq 0}$ – ланцюг Гарріса. Покладемо $A = B = \{B_1(0)\}$. З теореми 39 випливає, що $\mathbb{P}\{R_{X_0} < \infty\} = 1$. Оскільки $\inf\{h \geq 0 : X_h \in A\} < \infty\} \leq R_{X_0}$ для всіх можливих початкових станів, перша умова виконується. Для другої частини означення покладемо $q(x, y) := p := \mathbb{P}\{X_{h+1} = B_1(0) | X_h = B_1(0)\}$. З доведення теореми 39, більш точно з властивості (3.8), випливає, що $p > 0$.

Щоб перевірити рекурентність покладемо $\alpha = B_1(0)$. Тоді з теореми 39 маємо, що $R_\alpha < \infty$ майже напевно. Для аперіодичності достатньо помітити, що $\mathbb{P}\{X_1 = \alpha | X_0 = \alpha\} = p > 0$. \square

Теорема 47. Ланцюг Маркова $(X_h)_{h \geq 0}$ має стаціонарний розподіл I_∞ та

$$\lim_{n \rightarrow \infty} \|\mathbb{P}\{X_h = \cdot | X_0 = B_1(0)\} - I_\infty(\cdot)\|_{TV} = 0, \quad (3.9)$$

де $\|\cdot\|_{TV}$ позначає відстань повної варіації.

Доведення. З твердження 46 ми знаємо, що (X_h) аперіодичний рекурентний ланцюг Гарріса. З теореми 6.8.5 в [28, с. 323] цей ланцюг має стаціонарну міру. За теоремою 6.8.8 в [28, с. 324] аперіодичний рекурентний ланцюг Гарріса збігається до стаціонарного розподілу за повною варіацією як (3.9) за умови, що $R_{B_1(0)}$ майже напевно скінченний. Останнє забезпечено теоремою 39. \square

3.2.2 Граничні теореми для довжини найлівішого шляху. Використовуючи результати щодо поведінки ланцюга (X_h) , ми можемо проаналізувати

довжину найлівішого шляху VP-дерева, що побудоване на n випадкових точках. Введемо наступне випадкове блукання

$$S_0 := 0, \quad S_k := \sum_{h=0}^{k-1} G_h, \quad k \geq 1, \quad (3.10)$$

де, при фіксованому $(I_h)_{h \geq 0}$, $(G_h)_{h \geq 0}$ умовно незалежні випадкові величини такі, що розподіл G_h геометричний на \mathbb{N} з параметром $\frac{|I_h|}{2^d}$. Величина S_k дорівнює числу точок потрібних, щоб найлівіший шлях мав розмір хоча б k , оскільки нова вершина додається до нього, коли точка x_n потрапляє до області розміру $|I_h|$, що є, при фіксованому $(I_h)_{h \geq 0}$, випробуванням Бернуллі.

Розмір (кількість вершин) L_n найлівішого шляху в термінах випадкового блукання S_k може бути зображений так

$$L_n = \max\{k \geq 0 : S_k \leq n\}.$$

Довжина найлівішого шляху може бути легко підрахована з його розміру як $L_n - 1$.

Теорема 48. *Для випадкового VP-дерева розмір найлівішого шляху L_n задовольняє наступний слабкий закон великих чисел, при $n \rightarrow \infty$,*

$$\frac{L_n}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{-\log \tau}.$$

Зауважимо цікавий факт, що зазвичай вважається, що асимптотичні характеристики структур даних для задачі пошуку найближчого сусіда, включаючи VP-дерева, мають логарифмічний характер відносно кількості вершин n та експоненційну поведінку відносно розмірності простору d . На таку поведінку вказують численні результати математичних моделювань, зокрема роботи наведені під час огляду літератури, проте виявляється, що слабкий закон великих чисел для довжини найлівішого шляху не залежить від розмірності, проте вона все ще може фігурувати в наступних членах асимптотичного розкладу чи впливати на швидкість збіжності.

Щоб довести цю теорему нам спочатку необхідно довести дві допоміжні леми. Зауважимо, що якщо $X \sim \text{Geom}(a)$ та $Y \sim \text{Geom}(b)$, тоді

$$X \stackrel{st}{\leq} Y \Leftrightarrow a \geq b. \quad (3.11)$$

Лема 49. Для випадкових величин $G_k \sim \text{Geom}(\frac{|I_k|}{2^d})$ ($k \geq 0$) виконується

$$\frac{1}{k} \log G_k \xrightarrow{\mathbb{P}} -\log \tau,$$

при $k \rightarrow \infty$.

Доведення. Підрахуємо функцію розподілу $\frac{1}{k} \log G_k$. Зауважимо, що

$$\mathbb{P} \left\{ \frac{1}{k} \log G_k \leq x \right\} = \mathbb{E} \left\{ \mathbb{P} \left\{ \frac{1}{k} \log G_k \leq x \mid I_k \right\} \right\}, \quad (3.12)$$

та далі

$$\mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\} = \mathbb{P}\{G_k \leq \exp(kx) \mid I_k\} = 1 - (1 - |I_k|/2^d)^{\lfloor \exp(kx) \rfloor}. \quad (3.13)$$

З твердження 37 випливає

$$\tau^k \leq |I_k| \leq 2^d \tau^k.$$

Отже, можемо обмежити ймовірність в (3.13), підставляючи нижню та верхню оцінки $|I_k|$. Таким чином

$$1 - (1 - 2^{-d} \tau^k)^{\lfloor \exp(kx) \rfloor} \leq \mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\} \leq 1 - (1 - \tau^k)^{\lfloor \exp(kx) \rfloor}.$$

Обидві верхня та нижня оцінки збігаються, при $k \rightarrow \infty$, до

$$\begin{cases} 0, & \text{якщо } \log \tau + x < 0, \\ 1, & \text{якщо } \log \tau + x > 0. \end{cases}$$

Тому за теоремою про двох поліцейських $\mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\}$ збігається до того самого значення майже напевно. За теоремою Лебега про мажоровану збіжність, те саме виконується й для безумовної ймовірності $\mathbb{P} \left\{ \frac{1}{k} \log G_k \leq x \right\}$. Це завершує доведення. \square

Лема 50. Для випадкового блукання S_k , визначеного формулою (3.10), виконується

$$\frac{1}{k} \log S_k \xrightarrow{\mathbb{P}} -\log \tau$$

при $k \rightarrow \infty$.

Доведення. Запишемо функцію розподілу S_k

$$\mathbb{P}\{S_k \leq x\} = \mathbb{E}\{\mathbb{P}\{S_k \leq x | (I_h)_{0 \leq h < k}\}\}$$

та розглянемо більш детально умовну ймовірність під математичним сподіванням.

Оскільки $S_k \geq G_{k-1}$, отримуємо верхню оцінку

$$\mathbb{P}\{S_k \leq x | (I_h)_{0 \leq h < k}\} \leq \mathbb{P}\{G_{k-1} \leq x | (I_h)_{0 \leq h < k}\}, \quad (3.14)$$

та нерівність виконується й для математичного сподівання, тобто й для безумовної ймовірності.

Нагадаємо, що при фіксованому $(I_h)_{0 \leq h \leq k}$, випадкові величини $(G_h)_{0 \leq h \leq k}$ незалежні. Ми також знаємо, що $G_h \stackrel{st}{\leq} G_k$ ($0 \leq h < k$), оскільки послідовність $(|I_h|)_{h \geq 0}$ не зростає. Отже, ми можемо побудувати оцінку

$$\sum_{h=0}^{k-1} G_h \stackrel{st}{\leq} \sum_{h=0}^{k-1} G_k^{(h)} \stackrel{st}{\leq} k \max_{0 \leq h < k} G_k^{(h)} = k \hat{G}_k, \quad (3.15)$$

де $(G_k^{(h)})_{0 \leq h < k}$ – набір незалежних копій геометрично розподіленої випадкової величини з параметром $\tau^k/2^d$, що забезпечує $G_k \stackrel{st}{\leq} G_k^{(1)}$, та ми позначили $\hat{G}_k := \max_{0 \leq h < k} G_k^{(h)}$. Зауважимо, що нерівність справджується й для математичних сподівань, що значить, що вона виконується й для безумовних значень S_k .

Підрахуємо функцію розподілу \hat{G}_k

$$\mathbb{P}\left\{\frac{1}{k} \log \hat{G}_k \leq x\right\} = \mathbb{P}\{\hat{G}_k \leq \exp(kx)\} = \left(1 - (1 - \tau^k/2^d)^{\lfloor \exp(kx) \rfloor}\right)^k,$$

та звідси, при $k \rightarrow \infty$,

$$\mathbb{P}\left\{\frac{1}{k} \log \hat{G}_k \leq x\right\} \rightarrow \begin{cases} 0, & \text{якщо } \log \tau + x < 0, \\ 1, & \text{якщо } \log \tau + x \geq 0. \end{cases} \quad (3.16)$$

Тепер поєднаємо (3.14) та (3.15) й прологарифмуємо нерівності, щоб отримати

$$\log G_{k-1} \stackrel{st}{\leq} \log S_k \stackrel{st}{\leq} \log k + \log \hat{G}_k.$$

Поділивши все на k отримаємо

$$\frac{k-1}{k} \frac{\log G_{k-1}}{k-1} \stackrel{st}{\leq} \frac{\log S_k}{k} \stackrel{st}{\leq} \frac{\log k}{k} + \frac{\log \hat{G}_k}{k}.$$

З леми 49 та формули (3.16), спрямовуючи $k \rightarrow \infty$, маємо

$$\frac{1}{k} \log S_k \xrightarrow{d} -\log \tau.$$

Збіжність за ймовірністю випливає з того факту, що граничне значення є константою. \square

Доведення теореми 48. Почнемо з підрахунку функції розподілу $\frac{L_n}{\log n}$ в довільній точці $x > 0$ та $x \neq -1/\log \tau$

$$\begin{aligned} \mathbb{P} \left\{ \frac{L_n}{\log n} > x \right\} &= \mathbb{P} \{ L_n > x \log n \} = \mathbb{P} \{ S_{\lfloor x \log n \rfloor} \leq n \} \\ &= \mathbb{P} \{ \log S_{\lfloor x \log n \rfloor} \leq \log n \} = \mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{\log n} \leq 1 \right\} \\ &= \mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{x \log n} \leq \frac{1}{x} \right\}. \end{aligned}$$

З леми 50 отримуємо

$$\mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{x \log n} \leq \frac{1}{x} \right\} \rightarrow \begin{cases} 0, & 1/x < -\log \tau, \\ 1, & 1/x > -\log \tau, \end{cases}$$

при $k \rightarrow \infty$, та звідси

$$\mathbb{P} \left\{ \frac{L_n}{\log n} \leq x \right\} = 1 - \mathbb{P} \left\{ \frac{L_n}{\log n} > x \right\} \rightarrow \begin{cases} 1, & x > -1/\log \tau, \\ 0, & x < -1/\log \tau. \end{cases}$$

Оскільки збіжність за розподілом до константи означає й збіжність за ймовірністю, теорема доведена. \square

Зауважимо, що для доведення теореми 48 ми лише використовували, що $|I_h|/\tau^h$ відділені від нуля та від нескінченності деякими абсолютними константами. Наступні результати демонструють збіжність L_n вздовж підпоследовностей. Спочатку доведемо результат щодо випадкового блукання (S_k) . Нагадаємо, що I_∞ позначає стаціонарний розподіл ланцюга $(I_h/\tau^h)_{h \geq 0}$.

Лема 51. Для випадкового блукання S_k , що визначене в (3.10), при $k \rightarrow \infty$, виконується

$$\tau^k S_k \xrightarrow{d} G_\infty,$$

де G_∞ – випадковий ряд $\sum_{j=1}^{\infty} \xi_j$. Тут $(\xi_j)_{j \geq 0}$ позначають умовно незалежні, при фіксованому I_∞ , випадкові величини такі, що $\xi_j \sim \text{Exp}(\frac{|I_\infty|}{2^d \tau^j})$.

Доведення. Запишемо характеристичну функцію $\tau^k S_k$:

$$\begin{aligned} \mathbb{E} e^{it\tau^k S_k} &= \mathbb{E} \left[\mathbb{E} \left(e^{it\tau^k S_k} \mid (I_h)_{0 \leq h < k} \right) \right] = \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{e^{-it\tau^k} - (1 - |I_j|/2^d)} \right] \\ &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{1 - it\tau^k + O(\tau^{2k}) - (1 - |I_j|/2^d)} \right] \\ &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{|I_j|/2^d - it\tau^k + O(\tau^{2k})} \right] \\ &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{1}{1 - 2^d it\tau^{k-j} \frac{\tau^j}{|I_j|} + O(\frac{\tau^{2k}}{|I_j|})} \right] \\ &= \mathbb{E} \left[\prod_{j=1}^k \frac{1}{1 - 2^d it\tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\frac{\tau^k}{|I_{k-j}|})O(\tau^k)} \right] \\ &= \mathbb{E} \left[\prod_{j=1}^M \frac{1}{1 - 2^d it\tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k)} \cdot \prod_{j=M+1}^k \frac{1}{1 - 2^d it\tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k)} \right] \\ &= \mathbb{E} [A_{k,M}(t) \cdot B_{k,M}(t)], \end{aligned} \tag{3.17}$$

де $1 \leq M < k$ фіксований параметр та $O(\frac{\tau^k}{|I_{k-j}|}) = O(1)$ впливає з твердження 37.

З теореми 47 збіжність за повною варіацією означає й збіжність за розподілом. Крім того збіжність об'ємів також виконується, беручи до уваги особливості розглядуваних паралелепіпедів (див. формулу (3.4)). Отже, для кожного фіксованого M при $k \rightarrow \infty$ маємо

$$A_{k,M}(t) \xrightarrow{d} \prod_{j=1}^M \frac{1}{1 - 2^d it\tau^j / |I_\infty|}. \tag{3.18}$$

Тепер покажемо, що, для всіх $\varepsilon > 0$,

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \mathbb{P}\{|A_{k,M}(t)B_{k,M}(t) - A_{k,M}(t)| \geq \varepsilon\} = 0. \quad (3.19)$$

За нерівністю Маркова

$$\mathbb{P}\{|A_{k,M}(t)B_{k,M}(t) - A_{k,M}(t)| \geq \varepsilon\} \leq \frac{\mathbb{E}|A_{k,M}(t)B_{k,M}(t) - A_{k,M}(t)|}{\varepsilon}. \quad (3.20)$$

Спочатку покажемо, що $B_{k,M}(t)$ збігається до 1, показавши, що модуль та аргумент комплексного числа збігаються. Для модуля маємо

$$\begin{aligned} |B_{k,M}(t)|^2 &= \prod_{j=M+1}^k \frac{1}{(1 + O(\tau^k))^2 + (2^d t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k))^2}, \\ \log |B_{k,M}(t)|^2 &= - \sum_{j=M+1}^k \log \left[(1 + O(\tau^k))^2 + \left(2^d t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k) \right)^2 \right] \\ &= - \sum_{j=M+1}^k \log \left[1 + O(\tau^k) + 2^{2d} t^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2 + O(\tau^k) \right] \\ &= - \sum_{j=M+1}^k \left[2^{2d} t^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2 + O(\tau^k) + O(\tau^{4j}) + O(\tau^{2j+k}) + O(\tau^{2k}) \right] \\ &= - \sum_{j=M+1}^k [O(\tau^{2j}) + O(\tau^k)]. \end{aligned}$$

Розклад для логарифму виконується починаючи з деякого достатньо великого M . Ми також використали твердження 37, котре вказує, що $\tau^{k-j}/|I_{k-j}|$ обмежені невідповідними константами, що також не залежать від $k-j$. Зауважимо, що при фіксованому t , константи в залишках $O(\tau^{2j})$ та $O(\tau^k)$ рівномірні, тобто не залежать від j, k . Таким чином можемо записати

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \sum_{j=M+1}^k [O(\tau^{2j}) + O(\tau^k)] = 0 \quad \text{м.н.}$$

Для аргументу запишемо

$$\begin{aligned}
\arg B_{k,M}(t) &= - \sum_{j=M+1}^k \operatorname{Arg} \left(1 - 2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k) \right) \\
&= \sum_{j=M+1}^k \left[\frac{2^d t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|}}{\sqrt{1 + 2^{2d} t^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2}} + O(\tau^{3j}) + O(\tau^k) \right] \\
&= \sum_{j=M+1}^k [O(\tau^j) + O(\tau^k)].
\end{aligned}$$

Розклад для головного аргументу виконується починаючи з деякого достатньо великого M . Розмірковуючи як і до цього можна записати

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \arg B_{k,M}(t) = 0 \quad \text{м.н.}$$

Ми також показали в процесі, що починаючи з деяких достатньо великих k та M величина $|B_{k,M}(t)|$ обмежена не випадковою константою $\bar{B} > 0$ на довільному стохастичному шляху. Оскільки $|A_{k,M}(t)B_{k,M}(t)| = 1$, то $|A_{k,M}(t)|$ також обмежена деякою константою $\bar{A} > 0$. Отже, отримуємо

$$\frac{\mathbb{E}|A_{k,M}(t)||B_{k,M}(t) - 1|}{\varepsilon} \leq \frac{\bar{A}\mathbb{E}|B_{k,M}(t) - 1|}{\varepsilon}.$$

Ми вже показали, що

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} |B_{k,M}(t) - 1| = 0 \quad \text{м.н.}$$

Оскільки $|B_{k,M}(t) - 1| \leq 1 + \bar{B}$, то за теоремою Лебега про мажоровану збіжність

$$\lim_{M \rightarrow \infty} \lim_{k \rightarrow \infty} \mathbb{E}|B_{k,M}(t) - 1| = 0.$$

Поєднуючи всі частини разом, отримуємо (3.19).

Згідно з теоремою 64, наведеній в додатку, маємо, що, при $k \rightarrow \infty$,

$$e^{it\tau^k S_k} \xrightarrow{d} \lim_{M \rightarrow \infty} \prod_{j=1}^M \frac{1}{1 - 2^d i t \tau^j / |I_\infty|} = \prod_{j=1}^{\infty} \frac{1}{1 - 2^d i t \tau^j / |I_\infty|}.$$

Елементи послідовності $(e^{it\tau^k S_k})_{k \geq 0}$ рівномірно інтегровані як характеристичні функції, тому математичні сподівання теж збігаються, при $k \rightarrow \infty$,

$$\mathbb{E}e^{it\tau^k S_k} \rightarrow \mathbb{E} \prod_{j=1}^{\infty} \frac{1}{1 - 2^d i t \tau^j / |I_{\infty}|},$$

та граничне значення є характеристичною функцією G_{∞} , що завершує доведення. \square

Теорема 52. Для довільного фіксованого $T > 0$

$$L_{\lfloor T\tau^{-n} \rfloor} - n \xrightarrow{d} L_{\infty}^{(T)}$$

при $n \rightarrow \infty$ та функція розподілу $L_{\infty}^{(T)}$ рівна

$$\mathbb{P}\{L_{\infty}^{(T)} \leq m\} = 1 - \mathbb{P}\{G_{\infty} \leq T\tau^m\}$$

для всіх $m \in \mathbb{Z}$ з G_{∞} визначеному в лемі 51.

Доведення. Оскільки $L_{\lfloor T\tau^{-n} \rfloor} - n$ приймає значення лише в \mathbb{Z} ми досліджуємо функцію розподілу тільки на цілих значеннях. Для фіксованих $T > 0$ та $m \in \mathbb{Z}$ маємо

$$\begin{aligned} \mathbb{P}\{L_{\lfloor T\tau^{-n} \rfloor} - n > m\} &= \mathbb{P}\{L_{\lfloor T\tau^{-n} \rfloor} > n + m\} = \mathbb{P}\{S_{n+m} \leq \lfloor T\tau^{-n} \rfloor\} \\ &= \mathbb{P}\{\tau^{n+m} S_{n+m} \leq \tau^{n+m} \lfloor T\tau^{-n} \rfloor\}. \end{aligned}$$

З лемі 51 та того факту, що $\tau^n \lfloor T\tau^{-n} \rfloor \rightarrow T$ при $n \rightarrow \infty$, отримуємо

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\tau^{n+m} S_{n+m} \leq \tau^{n+m} \lfloor T\tau^{-n} \rfloor\} = \mathbb{P}\{G_{\infty} \leq T\tau^m\},$$

що завершує доведення. \square

3.3 Одновимірна збіжність

В цьому підрозділі ми аналізуємо випадкову рекурсію (3.3). Зафіксуємо довільний вимір $i \in \overline{1, d}$. Спрямовуючи $h \rightarrow \infty$ в (3.3) ми формально отримуємо стохастичне рівняння нерухомої точки для граничного значення X_{∞} послідовності $X_h^{(i)}$:

$$X_{\infty} \stackrel{d}{=} \min\{\tau^{-1}uX_{\infty}, 1\} + \min\{\tau^{-1}(1-u)X_{\infty}, 1\}, \quad (3.21)$$

де u рівномірно розподілена випадкова величина на $[0, 1]$. Покажемо, що $X_h^{(i)}$ збігається до X_∞ , при $h \rightarrow \infty$, використовуючи метод стискаючих відображень.

Теорема 53. Відображення $\mathcal{L}(X) \mapsto \mathcal{L}(f(X))$ в просторі ймовірнісних мір на $[1, 2]$, що наділений ℓ_1 -мінімальною метрикою, з

$$f(X) = \min \left\{ \frac{uX}{\tau}, 1 \right\} + \min \left\{ \frac{(1-u)X}{\tau}, 1 \right\}$$

є строго стискаючим, де u рівномірно розподілена випадкова величина на $[0, 1]$, незалежна від випадкової величини X .

Означення 54. ℓ_1 -мінімальна метрика між двома випадковими величинами X та Y задається формулою

$$\ell_1(X, Y) = \inf \mathbb{E}|\hat{X} - \hat{Y}|,$$

де інфімум береться по всіх парах випадкових величин (\hat{X}, \hat{Y}) таких, що $\hat{X} \sim X$ та $\hat{Y} \sim Y$.

Доведення теореми 53. Аналогічно до доведення твердження 37 можна перевірити, що якщо X приймає значення на $[1, 2]$, те ж саме виконується й для $f(X) = f(X, u)$.

Доведемо властивість стискання в два кроки.

Крок 1. Підрахуємо ℓ_1 -мінімальну відстань між $f(x, u_1)$ та $f(y, u_2)$ для фіксованих не випадкових чисел $x, y \in [1, 2]$. Маємо

$$\ell_1(f(x, u), f(y, u)) = \inf \mathbb{E}|f(x, u_1) - f(y, u_2)|.$$

Оскільки інфімум береться по всім можливим залежностям між u_1 та u_2 , можемо обрати $u_1 = u_2 = u$. Тоді

$$\ell_1(f(x, u_1), f(y, u_2)) \leq \int_{[0,1]} |f(x, u) - f(y, u)| du.$$

Припустимо на деякий час, що існує константа $\alpha \in (0, 1)$ така, що

$$\int_{[0,1]} |f(x, u) - f(y, u)| du \leq \alpha |x - y|. \quad (3.22)$$

Крок 2. Підрахуємо ℓ_1 -мінімальну відстань між $f(X, u_1)$ та $f(Y, u_2)$. Поєднуючи вищенаведені оцінки, отримуємо

$$\ell_1(f(X, u_1), f(Y, u_2)) \leq \alpha \mathbb{E}|\widehat{X} - \widehat{Y}|,$$

для довільної пари $(\widehat{X}, \widehat{Y})$ такої, що $\widehat{X} \sim X$ та $\widehat{Y} \sim Y$. Переходячи до нижньої межі по всім таким парами, отримуємо

$$\ell_1(f(X, u_1), f(Y, u_2)) \leq \alpha \ell_1(X, Y)$$

для деякої константи $\alpha \in (0, 1)$, що є означенням строго стискаючого відображення.

Залишається довести (3.22). Спочатку розділимо f на суму двох функцій.

$$f(x, u) = f_1(x, u) + f_2(x, u),$$

де

$$f_1(x, u) = \min \left\{ \frac{uX}{\tau}, 1 \right\} = \begin{cases} \frac{ux}{\tau}, & u < \frac{\tau}{x}, \\ 1, & u \geq \frac{\tau}{x} \end{cases}$$

та $f_2(x, u) = \min \left\{ \frac{(1-u)X}{\tau}, 1 \right\}$.

Підрахуємо наступний інтеграл за умови $x \leq y$:

$$\begin{aligned} \int_0^1 |f_1(x, u) - f_1(y, u)| du &= \int_0^{\tau/y} \left(\frac{uy}{\tau} - \frac{ux}{\tau} \right) du \\ &\quad + \int_{\tau/y}^{\tau/x} \left(1 - \frac{ux}{\tau} \right) du + \int_{\tau/x}^1 (1 - 1) du \\ &= \int_0^{\tau/y} \frac{u}{\tau} (y - x) du + \int_{\tau/y}^{\tau/x} du - \int_{\tau/y}^{\tau/x} \frac{x}{\tau} du \\ &= \frac{y - x}{\tau} \frac{u^2}{2} \Big|_0^{\tau/y} + \left(\frac{\tau}{x} - \frac{\tau}{y} \right) - \frac{x}{\tau} \frac{u^2}{2} \Big|_{\tau/y}^{\tau/x} \\ &= \frac{y - x}{2\tau} \frac{\tau^2}{y^2} + \left(\frac{\tau}{x} - \frac{\tau}{y} \right) - \frac{x}{2\tau} \left(\frac{\tau^2}{x^2} - \frac{\tau^2}{y^2} \right) \\ &= \frac{\tau(y - x)}{2y^2} + \frac{\tau}{x} - \frac{\tau}{y} - \frac{\tau}{2x} + \frac{\tau x}{2y^2} \\ &= \frac{\tau x(y - x) + 2\tau y^2 - 2\tau xy - \tau y^2 + \tau x^2}{2xy^2} \\ &= \frac{-\tau xy + \tau y^2}{2xy} = \frac{\tau}{2} \frac{y - x}{xy}. \end{aligned}$$

Аналогічно, якщо $y \leq x$, то отримуємо

$$\int_0^1 |f_1(x, u) - f_1(y, u)| du \leq \frac{\tau}{2} \frac{x - y}{xy},$$

що призводить до

$$\int_0^1 |f_1(x, u) - f_1(y, u)| du \leq \frac{\tau}{2} \frac{|x - y|}{xy}. \quad (3.23)$$

Тепер оцінимо той самий інтеграл для f_2 , де

$$f_2(x, u) = \begin{cases} 1, & 1 - u \geq \frac{\tau}{x}, \\ \frac{(1-u)x}{\tau}, & 1 - u < \frac{\tau}{x}. \end{cases}$$

Під інтегралом зробимо заміну $v = 1 - u$:

$$\int_0^1 |f_2(x, u) - f_2(y, u)| du = \int_0^1 |f_2(x, 1 - v) - f_2(y, 1 - v)| dv,$$

де миттєво отримуємо ту саму оцінку, що й (3.23), оскільки $f_2(x, 1 - v) = f_1(x, v)$.

Нарешті,

$$\begin{aligned} \int_{[0,1]} |f(x, u) - f(y, u)| du &\leq \int_0^1 |f_1(x, u) - f_1(y, u)| du \\ &\quad + \int_0^1 |f_2(x, u) - f_2(y, u)| du \\ &\leq \frac{\tau}{2} \frac{|x - y|}{xy} + \frac{\tau}{2} \frac{|x - y|}{xy} = \tau \frac{|x - y|}{xy} \leq \tau |x - y|, \end{aligned}$$

що доводить (3.22) з $\alpha = \tau$. □

Наслідок 55. *Стохастичне рівняння нерухомої точки (3.21) має єдиний розв'язок X_∞ в просторі ймовірнісних мір на $[1, 2]$ та $X_h^{(i)} \xrightarrow{d} X_\infty$, при $h \rightarrow \infty$, для кожного фіксованого $i = 1, \dots, d$.*

Зауваження 56. Ми отримали той самий результат, що й у попередньому розділі без звертання до геометричної інтерпретації й аналізу, проте з явним використанням розподілу послідовності точок x_1, x_2, \dots . Це також дозволило нам вивести стохастичне рівняння нерухомої точки для X_∞ . Нагадаємо, що

I_∞ позначає d -вимірний прямокутний паралелепіпед зі сторонами паралельними осям координат, його розміри по всім розмірностям можуть бути задані набором $(X_\infty^{(1)}, \dots, X_\infty^{(d)})$, де $(X_\infty^{(i)})_{1 \leq i \leq d}$ незалежні копії в.в. X_∞ .

Виявляється, що розподіл X_∞ можна знайти явно, використовуючи стохастичне рівняння нерухомої точки (3.21) шляхом прямого обчислення її функції розподілу вигляду¹ $F(t) = \mathbb{P}\{X_\infty < t\}$. Розв'язок суттєво залежить від параметра τ .

Розглянемо спочатку випадок $\tau \leq 1/2$.

Для $t \in [1, 2]$ запишемо

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= \mathbb{P}\left\{\min\left\{\frac{uX_\infty}{\tau}, 1\right\} + \min\left\{\frac{(1-u)X_\infty}{\tau}, 1\right\} < t\right\} \\ &= \mathbb{P}\left\{1 + 1 < t, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &\quad + \mathbb{P}\left\{\frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1\right\} \\ &\quad + \mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &\quad + \mathbb{P}\left\{1 + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} < 1\right\}. \end{aligned}$$

Перший доданок очевидно рівний 0, оскільки ми припускаємо, що $t \leq 2$. Другий доданок також рівний 0, оскільки події $\{\frac{uX_\infty}{\tau} < 1\}$ та $\{\frac{(1-u)X_\infty}{\tau} < 1\}$ не можуть відбутись одночасно при $\tau \leq 1/2$. Оскільки останні два доданки симетричні, то

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= 2\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= 2\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}, u \leq \frac{X_\infty - \tau}{X_\infty}\right\} = 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}\right\}. \end{aligned}$$

Тут в другому переході ми використали, що подія $\{\frac{uX_\infty}{\tau} + 1 < t\}$ є наслідком події $\{\frac{uX_\infty}{\tau} < 1\}$. Та в останньому переході було використано, що подія $\{u <$

¹Звернемо увагу, що в цьому вигляді використовується знак строго менше.

$\frac{\tau(t-1)}{X_\infty}$ } включає в себе подію $\{u \leq \frac{X_\infty - \tau}{X_\infty}\}$, оскільки $\tau t < 1 \leq X_\infty$. Нарешті,

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}\right\} = 2\mathbb{E}\left(\mathbb{P}\left\{\frac{\tau(t-1)}{X_\infty} > u\right\} \mid X_\infty\right) \\ &= 2\mathbb{E}\left(\frac{\tau(t-1)}{X_\infty}\right) = 2\tau\mathbb{E}\frac{1}{X_\infty}(t-1). \end{aligned} \quad (3.24)$$

Також маємо особливий випадок $X_\infty = 2$:

$$\begin{aligned} \mathbb{P}\{X_\infty = 2\} &= \mathbb{P}\left\{1 + 1 = 2, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= \mathbb{P}\left\{u \geq \frac{\tau}{X_\infty}, u \leq 1 - \frac{\tau}{X_\infty}\right\} = 1 - 2\tau\mathbb{E}\frac{1}{X_\infty}. \end{aligned} \quad (3.25)$$

З (3.24) та (3.25) можна бачити, що X_∞ є комбінацією рівномірного розподілу на $[1, 2)$ та атому в точці 2. Нам лише залишається підрахувати константу $\mathbb{E}\frac{1}{X_\infty}$.

$$\begin{aligned} \mathbb{E}\frac{1}{X_\infty} &= \frac{1}{2}\mathbb{P}\{X_\infty = 2\} + \int_1^2 \frac{1}{t} d\mathbb{P}\{X_\infty < t\} \\ &= \frac{1}{2}\left(1 - 2\tau\mathbb{E}\frac{1}{X_\infty}\right) + \int_1^2 \frac{2\tau\mathbb{E}\frac{1}{X_\infty}}{t} dt \\ &= \frac{1}{2}\left(1 - 2\tau\mathbb{E}\frac{1}{X_\infty}\right) + 2\tau\mathbb{E}\frac{1}{X_\infty}(\log 2 - \log 1). \end{aligned} \quad (3.26)$$

Розв'язавши (3.26) для невідомої величини $\mathbb{E}\frac{1}{X_\infty}$, отримуємо, що

$$\mathbb{E}\frac{1}{X_\infty} = \frac{1}{2\tau(1 - 2\log 2 + 1/\tau)}.$$

Відтепер припускаємо, що $\tau > 1/2$. Почнемо, як і в попередньому випадку, проте запишемо всі варіанти для мінімумів окремо.

Ситуація $X_\infty = 2$ можлива лише коли обидва мінімуми набувають значе-

ння 1:

$$\begin{aligned}
\mathbb{P}\{X_\infty = 2\} &= \mathbb{P}\left\{1 + 1 = 2, 1 \leq \frac{uX_\infty}{\tau}, 1 \leq \frac{(1-u)X_\infty}{\tau}\right\} \\
&= \mathbb{P}\left\{\frac{\tau}{X_\infty} \leq u \leq \frac{X_\infty - \tau}{X_\infty}, X_\infty \geq 2\tau\right\} \\
&= \int_{[2\tau, 2]} \left(1 - \frac{\tau}{x}\right) dF(x) - \int_{[2\tau, 2]} \frac{\tau}{x} dF(x) \\
&= 1 - F(2\tau) - 2\tau \int_{[2\tau, 2]} \frac{1}{x} dF(x) \\
&= 1 - F(2\tau) - 2\tau \left(\mathbb{E} \frac{1}{X_\infty} - \int_{[1, 2\tau]} \frac{1}{x} dF(x)\right).
\end{aligned} \tag{3.27}$$

Коли $X_\infty < 2$ існують дві можливості, що лише один із мінімумів в (3.21) рівний 1, та обидві ці ситуації симетричні, отож ми розпишемо лише одну із них:

$$\begin{aligned}
&\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\
&= \mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}, X_\infty \geq \tau t\right\} + \mathbb{P}\left\{u \leq \frac{X_\infty - \tau}{X_\infty}, X_\infty < \tau t\right\} \\
&= \int_{[\tau t, 2]} \frac{\tau(t-1)}{x} dF(x) + \int_{[1, \tau t]} \left(1 - \frac{\tau}{x}\right) dF(x) \\
&= \tau(t-1) \left(\mathbb{E} \frac{1}{X_\infty} - \int_{[1, \tau t]} \frac{1}{x} dF(x)\right) + F(\tau t) - \tau \int_{[1, \tau t]} \frac{1}{x} dF(x).
\end{aligned} \tag{3.28}$$

Остання можливість, що $X_\infty < 2$, трапляється, коли обидва мінімуми менші за 1.

$$\begin{aligned}
&\mathbb{P}\left\{\frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1\right\} \\
&= \mathbb{P}\left\{\frac{X_\infty - \tau}{X_\infty} < u < \frac{\tau}{X_\infty}, x < \tau t\right\} = \int_{[1, \tau t]} \frac{\tau}{x} dF(x) - \int_{[1, \tau t]} \left(1 - \frac{\tau}{x}\right) dF(x) \\
&= 2\tau \int_{[1, \tau t]} \frac{1}{x} dF(x) - F(\tau t).
\end{aligned} \tag{3.29}$$

Поєднавши (3.28) та (3.29), отримуємо, при $t \in [1, 2]$,

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= \mathbb{P}\left\{\frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1\right\} \\ &\quad + 2\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= F(\tau t) + 2\tau(t-1) \left(\mathbb{E}\frac{1}{X_\infty} - \int_{[1, \tau t)} \frac{1}{x} dF(x) \right). \end{aligned} \tag{3.30}$$

Тепер покажемо як отримати точні значення ймовірності (3.30). На кроці 1 припустимо, що $t \in [1, 1/\tau]$, тоді $\tau t < 1$ й з (3.30):

$$\mathbb{P}\{X_\infty < t\} = 0 + 2\tau(t-1) \left(\mathbb{E}\frac{1}{X_\infty} - 0 \right) = 2\tau(t-1)\mathbb{E}\frac{1}{X_\infty}.$$

Також,

$$dF(t) = 2\tau\mathbb{E}\frac{1}{X_\infty}dt.$$

Ми повторюємо процедуру підрахунку ймовірностей та похідних для $t \in [1/\tau, 1/\tau^2]$ й так далі на наступних інтервалах, використовуючи результати з попередніх інтервалів. Більш загально, підрахуємо значення на інтервалі на кроці k , тобто коли $t \in [1/\tau^{k-1}, 1/\tau^k]$.

Маємо $\tau t \in [1/\tau^{k-2}, 1/\tau^{k-1}]$, що означає, що $F(\tau t)$ є значенням з попереднього інтервалу й ми його вже можемо використовувати. Також маємо інтегральну частину в (3.30):

$$\int_{[1, \tau t)} \frac{1}{x} dF(x) = \int_1^{1/\tau} \frac{1}{x} dF(x) + \int_{1/\tau}^{1/\tau^2} \frac{1}{x} dF(x) + \dots + \int_{1/\tau^{k-2}}^{\tau t} \frac{1}{x} dF(x),$$

де інтеграли беруться, використовуючи підраховані похідні на перших $k-1$ інтервалах відповідно.

На деякому k матимемо $1/\tau^k > 2$, тому ми припиняємо підрахунок інтервалів. У нас також є лише одна невідома величина $\mathbb{E}\frac{1}{X_\infty}$. Ми підраховуємо це значення явно, використовуючи знайдені ймовірності, та розв'язуємо рівняння відносно цієї величини.

Приклад 57. Знайдемо X_∞ , коли $1/\tau < 2 \leq 1/\tau^2$ ($k=2$), тобто $1/2 < \tau \leq 1/\sqrt{2}$.

Якщо $t \leq 1/\tau$, отримуємо

$$\mathbb{P}\{X_\infty < t\} = 2\tau(t-1)\mathbb{E}\frac{1}{X_\infty}, \quad dF(t) = 2\tau\mathbb{E}\frac{1}{X_\infty}dt. \quad (3.31)$$

Якщо $t \geq 1/\tau$,

$$\int_{[1,\tau t]} \frac{1}{x}dF(x) = \int_1^{\tau t} \frac{2\tau\mathbb{E}\frac{1}{X_\infty}}{x}dx = 2\tau\mathbb{E}\frac{1}{X_\infty}\log(\tau t). \quad (3.32)$$

Після підстановки (3.31) та (3.32) в (3.30), отримуємо

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= F(\tau t) + 2\tau(t-1) \left(\mathbb{E}\frac{1}{X_\infty} - 2\tau\mathbb{E}\frac{1}{X_\infty}\log(\tau t) \right) = \\ &= 2\tau(\tau t - 1)\mathbb{E}\frac{1}{X_\infty} + 2\tau(t-1)\mathbb{E}\frac{1}{X_\infty} - 4\tau^2(t-1)\log(\tau t) = \\ &= 2\tau\mathbb{E}\frac{1}{X_\infty}(t + \tau t - 2\tau t\log(\tau t) + 2\tau\log(\tau t) - 2), \end{aligned}$$

та диференціал рівний

$$dF(t) = 2\tau\mathbb{E}\frac{1}{X_\infty} \left(1 + \tau - 2\tau(1 + \log(\tau t)) + 2\tau\frac{1}{t} \right) dt.$$

Для підрахунку $\mathbb{P}\{X_\infty = 2\}$ скористаємось формулою (3.27).

$$\begin{aligned} \mathbb{P}\{X_\infty = 2\} &= 1 - F(2\tau) - 2\tau \left(\mathbb{E}\frac{1}{X_\infty} - \int_{[1,2\tau]} \frac{1}{x}dF(x) \right) = 1 - \mathbb{P}\{X_\infty < 2\} \\ &= 1 - 2\tau\mathbb{E}\frac{1}{X_\infty}(2 + 2\tau - 4\tau\log(2\tau) + 2\tau\log(2\tau) - 2) \\ &= 1 - 4\tau^2\mathbb{E}\frac{1}{X_\infty}(1 - \log(2\tau)). \end{aligned}$$

Залишається знайти константу $\mathbb{E}\frac{1}{X_\infty}$, що робиться так:

$$\begin{aligned}
\mathbb{E}\frac{1}{X_\infty} &= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E}\frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E}\frac{1}{X_\infty}}{x} dx \\
&\quad + \int_{1/\tau}^2 \frac{2\tau \mathbb{E}\frac{1}{X_\infty} (1 + \tau - 2\tau(1 + \log(\tau t)) + 2\tau \frac{1}{t})}{x} dx \\
&= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E}\frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E}\frac{1}{X_\infty}}{x} dx \\
&\quad + 2\tau \mathbb{E}\frac{1}{X_\infty} \left(\int_{1/\tau}^2 \frac{1 + \tau - 2\tau}{x} dx + \int_{1/\tau}^2 \frac{-2\tau \log \tau x}{x} dx + \int_{1/\tau}^2 \frac{2\tau}{x^2} dx \right) \\
&= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E}\frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E}\frac{1}{X_\infty}}{x} dx + 2\tau \mathbb{E}\frac{1}{X_\infty} * \\
&\quad * \left((1 - \tau)(\log 2 - \log(1/\tau)) - \tau(\log^2(2\tau) - \log^2(\tau/\tau)) - 2\tau \left(\frac{1}{2} - \tau \right) \right) \\
&= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E}\frac{1}{X_\infty} (1 - \log(2\tau)) \right) - 2\tau \mathbb{E}\frac{1}{X_\infty} \log(\tau) \\
&\quad + 2\tau \mathbb{E}\frac{1}{X_\infty} \left((1 - \tau) \log(2\tau) - \tau \log^2(2\tau) + 2\tau \left(\tau - \frac{1}{2} \right) \right).
\end{aligned}$$

Розв'язавши це рівняння відносно невідомої величини $\mathbb{E}\frac{1}{X_\infty}$, отримуємо:

$$\mathbb{E}\frac{1}{X_\infty} = \frac{1}{2(1 - 4\tau^3 + 4\tau^2 + 2\tau^2 \log^2(2\tau) + 2\tau \log(\tau) - 2\tau \log(2\tau))}.$$

3.4 Балансування

В цьому підрозділі ми розглядатимемо лише частковий випадок $d = 1$. У цьому випадку $|I_h| = I_h^{(1)}$. Надалі для простоти позначень ми не будемо писати верхній індекс (1) у величинах, наприклад, писатимемо I_h замість $I_h^{(1)}$.

Нагадаємо, що бінарне дерево з n вершин називається збалансованим, якщо його висота рівна $\lceil \log(n+1)/\log 2 \rceil - 1$, що є її абсолютним мінімумом. Процедурою балансування в комп'ютерних науках називають довільні операції над деревом, що призводять до зменшення його висоти та не порушують його характеристичні властивості.

Підрахуємо математичне сподівання I_h – розміру розбиттів простору, що породжені найлівишим шляхом випадкового VP-дерева. Для цього скористаємось рекурсивною формулою (3.2):

$$\mathbb{E}I_h = \mathbb{E} \left[\min\{u_h I_{h-1}, \tau^h\} + \min\{(1 - u_h)I_{h-1}, \tau^h\} \right].$$

З адитивності математичного сподівання та оскільки u_h й $1 - u_h$ мають однаковий рівномірний розподіл на $[0, 1]$, середні значення обох мінімумів будуть рівними. Звідси

$$\begin{aligned} \mathbb{E}I_h &= 2\mathbb{E} \min\{u_h I_{h-1}, \tau^h\} = 2\mathbb{E} \left[\tau^h + \min\{u_h I_{h-1} - \tau^h, 0\} \right] \\ &= 2\tau^h + 2\mathbb{E}(u_h I_{h-1} - \tau^h) \mathbb{1}_{\{u_h I_{h-1} - \tau^h \leq 0\}} \\ &= 2\tau^h + 2\mathbb{E} \left[(u_h I_{h-1} - \tau^h) \mathbb{1}_{\{u_h I_{h-1} - \tau^h \leq 0\}} \middle| I_{h-1} \right] \\ &= 2\tau^h + 2\mathbb{E} \left[u_h I_{h-1} \mathbb{1}_{\{u_h \leq \tau^h / I_{h-1}\}} \middle| I_{h-1} \right] - 2\tau^h \mathbb{E} \left[\mathbb{1}_{\{u_h \leq \tau^h / I_{h-1}\}} \middle| I_{h-1} \right] \\ &= 2\tau^h + 2\mathbb{E} \left[\frac{\tau^{2h}}{2I_{h-1}^2} I_{h-1} \middle| I_{h-1} \right] - 2\tau^h \mathbb{E} \left[\frac{\tau^h}{I_{h-1}} \middle| I_{h-1} \right] \\ &= 2\tau^h + \tau^{2h} \mathbb{E} \frac{1}{I_{h-1}} - 2\tau^{2h} \mathbb{E} \frac{1}{I_{h-1}} \\ &= 2\tau^h - \tau^{2h} \mathbb{E} \frac{1}{I_{h-1}}. \end{aligned} \tag{3.33}$$

В процесі ми також скористались твердженням 37, звідки випливає, що $\tau^h / I_{h-1} \leq \tau < 1$. В термінах $X_h = I_h / \tau^h$ рівняння (3.33) приймає дуже простого вигляду

$$\mathbb{E}X_h = 2 - \tau \mathbb{E} \frac{1}{X_{h-1}}.$$

Ідея балансування випадкового VP-дерева полягає у тому, щоб обрати параметр τ так, щоб у середньому підпростори ділились навпіл, таким чином кількість точок, що потраплятимуть в ліві на праві піддерева будуть рівними в середньому. Для цього покладемо

$$\mathbb{E}I_1 = 1,$$

оскільки початковий простір має розмір 2. З рівняння (3.33) маємо

$$1 = 2\tau - \tau^2 \frac{1}{2} \tag{3.34}$$

для невідомого τ . Розв'язками цього квадратного рівняння є значення $2 - \sqrt{2}$ та $2 + \sqrt{2}$, проте лише перше з них належить інтервалу $(0, 1)$. Остаточо ми можемо сформулювати гіпотезу щодо балансування.

Гіпотеза 58. *Випадкове VP-дерево при $d = 1$ буде найбільш збалансованим при значенні параметра $\tau = 2 - \sqrt{2}$.*

Емпірична перевірка цієї гіпотези буде виконана в підрозділі математичного моделювання.

3.5 Математичне моделювання

На початку наведемо декілька зображень випадкових VP-дерев за різних значень параметра τ , щоб мати уявлення про їх структуру та про вплив цього параметра. Нагадаємо, що в роботі було наведено схему вибору порогового значення як τ^h , де h – або рівень вершини плюс один, це дерево позначатиметься як звичайне VP-дерево (VPTree), або як кількість “поворотів вліво” на шляху від кореня до вершини плюс один, це дерево позначатиметься як VPTreeLefts. В обох випадках генерувалось $n = 1000$ точок в $[-1, 1]$, тобто при $d = 1$.

На рисунку 3.2 зображено приклади реалізацій звичайних VP-дерев. На рисунку 3.3 зображено реалізації дерев VPTreeLefts повторюючи деякі значення параметра. Можна бачити закономірність, що друга модель “зміщена вліво” за рахунок зменшення величини h , й загальна тенденція полягає в зміщенні з права наліво при зростанні параметра τ . Нагадаємо, що розподіл найлівішого шляху в обох випадках однаковий.

До того ж на рисунку 3.2 можна візуально побачити, що гіпотеза 58 справджується для звичайного VP-дерева. Щоб перевірити це більш детально згенеруємо для різних значень параметра τ по 1000 незалежних випадкових VP-дерев й підрахуємо їх середню висоту. Результати цього представлені в таблиці 3.1. Дійсно, найменша середня висота досягається при значення

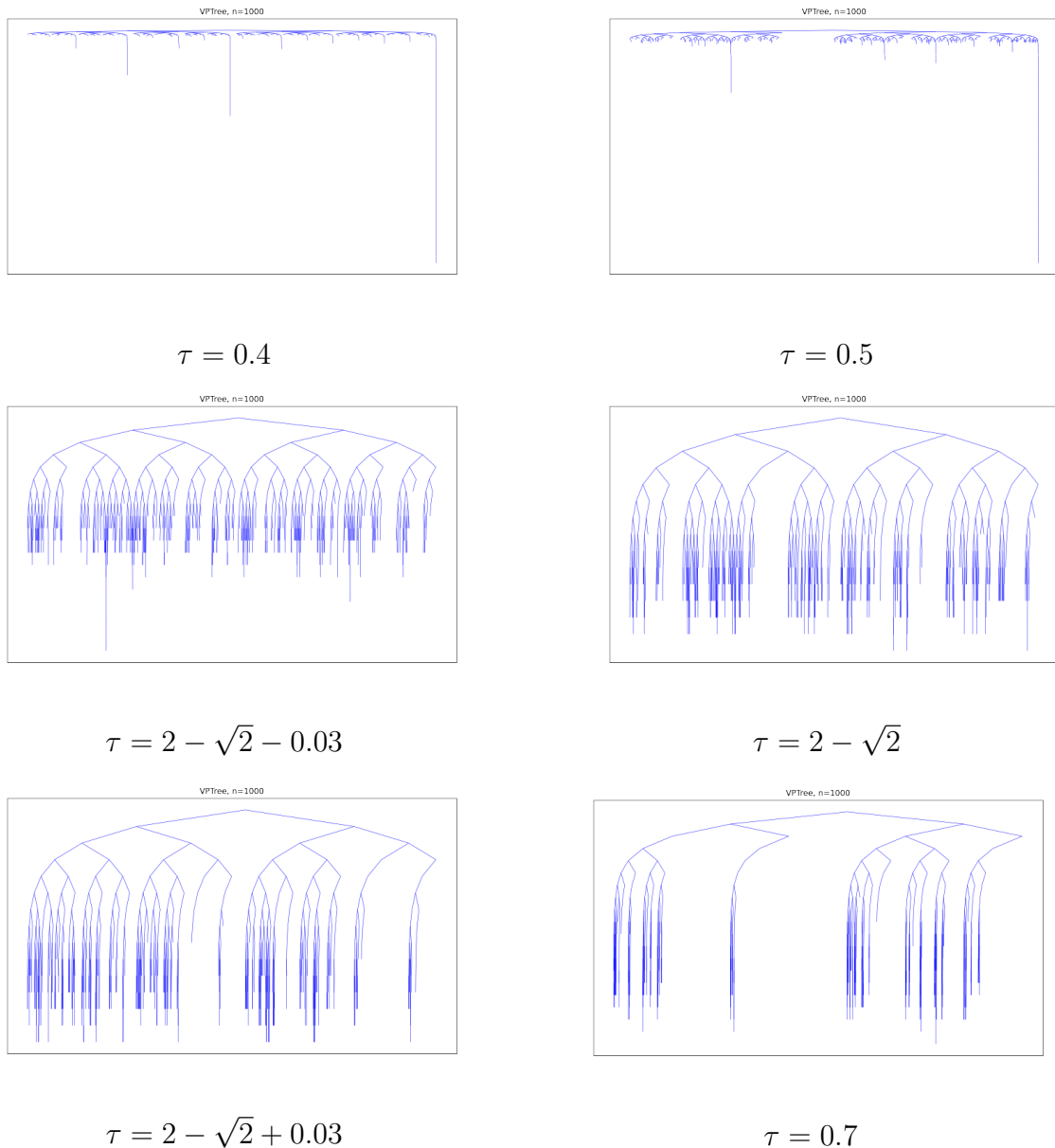


Рис. 3.2: Приклади реалізації випадкового VP-дерева за різних значень параметра τ .

$\tau = 2 - \sqrt{2} \approx 0.585786 \approx 0.59$, що підтверджує висунуту гіпотезу. Зауважимо, що оскільки $n = 1000$, то повністю збалансоване дерево мало б висоту рівну 9.

В підрозділі 3.3 було підраховано граничний розподіл для нормованої довжини розбиттів, що породжені найлівішим шляхом, в одновимірному випадку. На наступних графіках можна переконатись в коректності отриманих теоретичних результатів. На рисунку 3.4 зображено одночасні графіки отриманих теоретичних значень $F(t)$ та статистичних значень. Під статистичними зна-

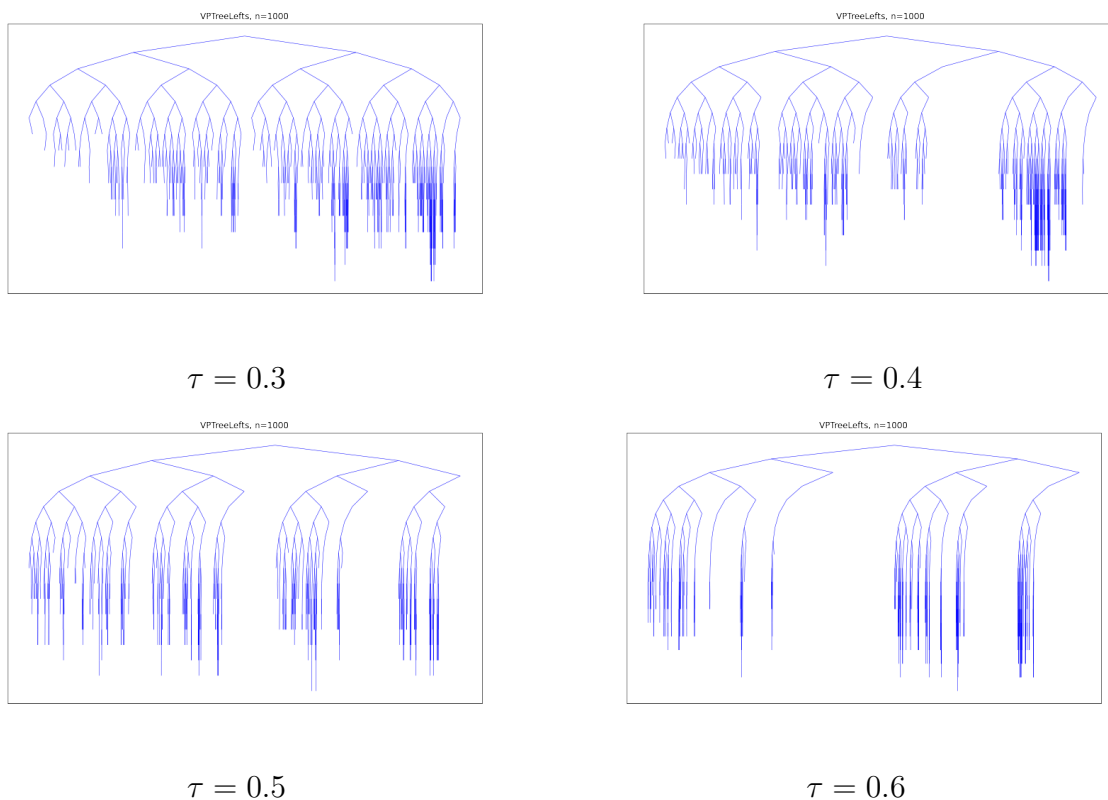


Рис. 3.3: Приклади реалізації дерева VPTreeLefts за різних значень параметра τ .

τ	Середня висота	τ	Середня висота
0.55	25.774	0.62	15.874
0.56	19.529	0.63	16.257
0.57	16.455	0.64	16.63
0.58	15.042	0.65	17.044
0.59	14.919	0.66	17.521
0.6	15.158	0.67	18.052
0.61	15.475	0.68	18.518

Табл. 3.1: Середня висота звичайного випадкового VP-дерева при різних значеннях параметра τ .

ченнями ми розуміємо емпіричну кумулятивну функцію розподілу, що отримана зі значень $X_h^{(i)}$, що були отримані з 10000 послідовних ітерацій процесу (3.3).

На рисунку 3.5 зображено графіки теоретичних та статистичних значень $F(t)$ при τ , що належить інтервалу з прикладу 57.

Завершуючи зображення функції розподілу, без підрахунку та вказування

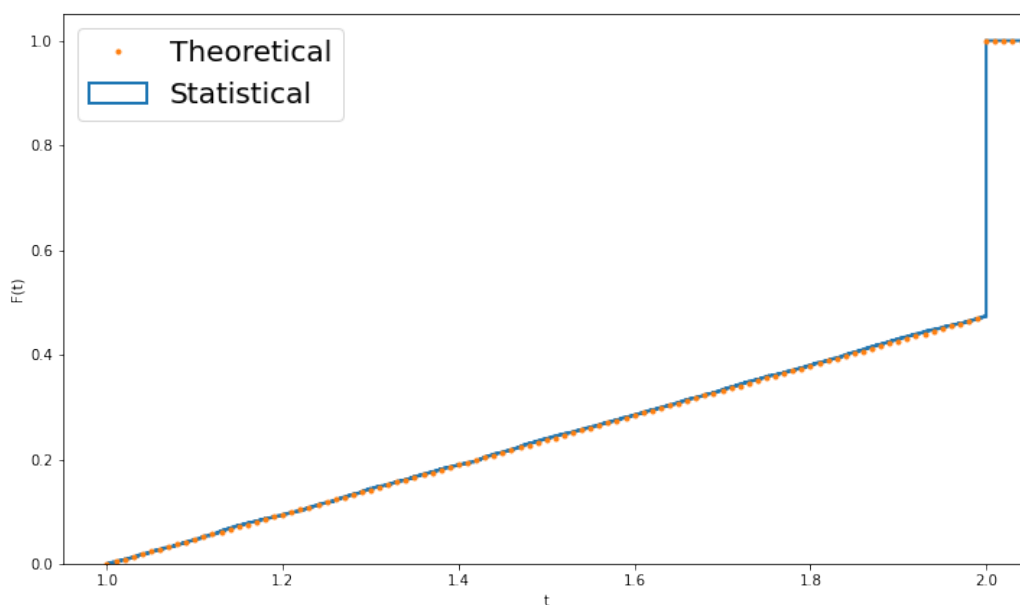


Рис. 3.4: Графік теоретичних та статистичних значень функції розподілу $F(t)$ в.в. X_∞ при $\tau = 0.4$.

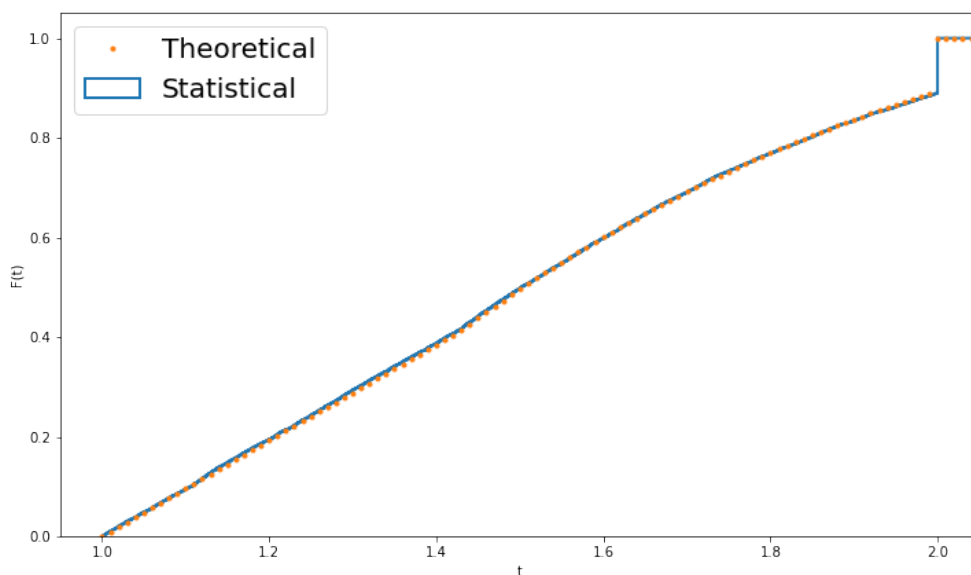
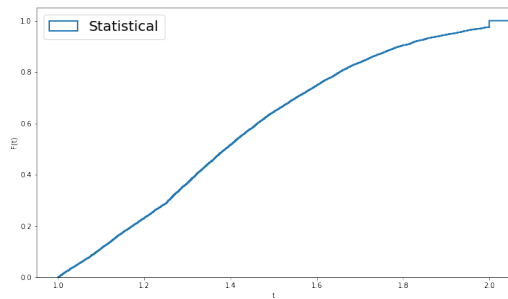
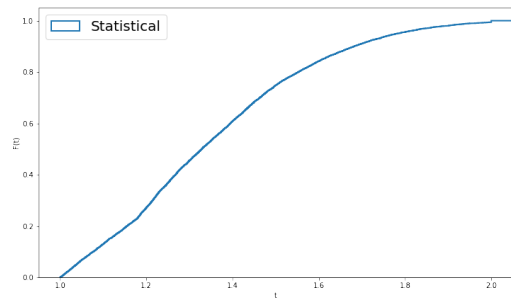


Рис. 3.5: Графік теоретичних та статистичних значень функції розподілу $F(t)$ в.в. X_∞ при $\tau = 0.7$.

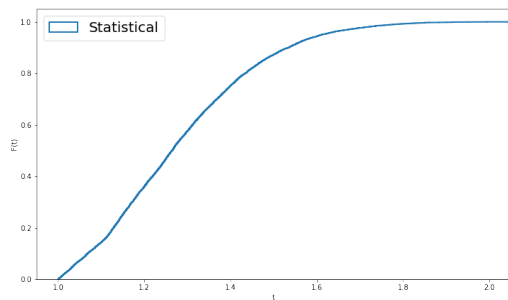
її теоретичних значень ми наводимо приклад як вона в подальшому змінюється з ростом τ зображаючи лише її статистичні значення на рисунку 3.6



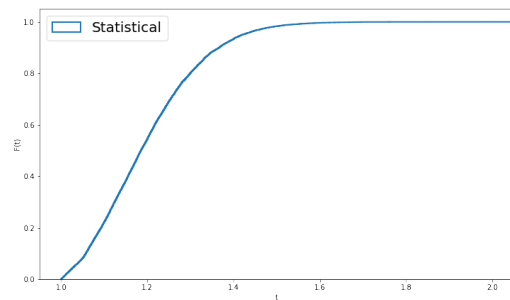
$$\tau = 0.8$$



$$\tau = 0.85$$



$$\tau = 0.9$$



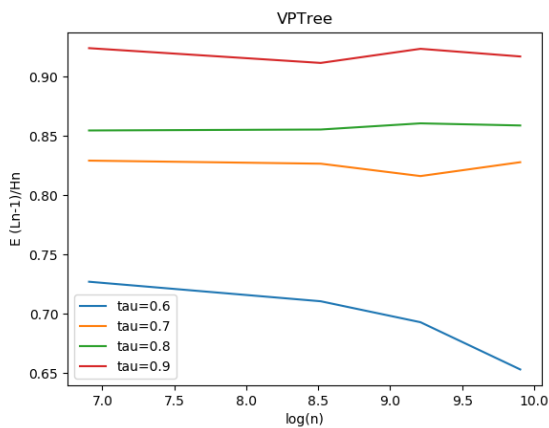
$$\tau = 0.95$$

Рис. 3.6: Еволюція підрахованої статистично функції розподілу $F(t)$ в.в. X_∞ при збільшенні значень τ .

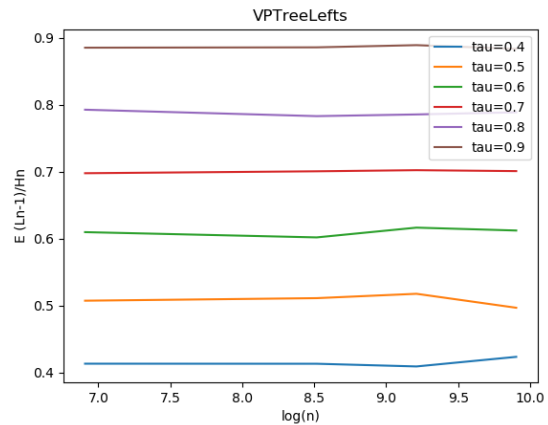
при різних значеннях параметра. Зауважимо, що з графіків випливає, що при $\tau \rightarrow 1$ виконується $X_\infty \xrightarrow{d} 1$. Проте цей результат є досить природнім, оскільки при $\tau = 1$ послідовність $X_h^{(i)}$ на кожному кроці може лише спадати та з твердження 37, що виконується й при $\tau = 1$, обмежена знизу величиною 1. Ця властивість більш строго була досліджена в роботах [1, 45].

Нарешті наведемо результат, що отриманий суто шляхом моделювання та який демонструє зв'язок між таким неklasичним функціоналом як довжина найлівішого шляху та класичною характеристикою – висотою дерева. На рисунку 3.7 зображено середнє відношень довжини найлівішого шляху та загальної висоти дерева, позначеної H_n , за різних параметрів τ та різної кількості вершин у дереві n ($d = 1$), усереднення відбувалось з 50 незалежних однакових генерацій дерева.

З графіків випливає, зокрема для другого правила вибору порогового значення, залежність між дослідженим функціоналом та таким класичним фун-



Звичайне VP-дерево



VP-дерево з модифікованим правилом вибору порогового значення

Рис. 3.7: Графік середнього відношення довжини найлівішого шляху до висоти дерева за різних параметрів τ та кількості вершин n .

кціоном як висота, що слугує мотивацією для подальших досліджень з використанням вже отриманих результатів.

3.5.1 Програмний код мовою Python3. На початку наведемо імплементацію VP-дерева. В лістингу 3.1 наведено загальний клас дерева, з означеними спільними функціоналами дерева як розмір найлівішого шляху `get_left_path()` та висота `get_height()`, що реалізовані звичайним проходом по цьому шляху та проходом дерева в глибину відповідно. В лістингу 3.2 наведено реалізації звичайного VP-дерева та модифікованого `VPTreeLefts`. Відрізняються вони саме реалізацією функції додавання нової точки, точніше способом підрахунку величини h .

```

1 # tree.py
2 import abc
3
4
5 class Tree(abc.ABC):
6     # abstract tree class
7
8     def __init__(self):
9         self.parent = None # parent of each node
10        self.sons = [] # list of sons
11
12    @abc.abstractmethod

```

```

13  def add(self, points):
14      # adding a new point to the tree
15      pass
16
17  def get_height(self):
18      # return a height of the tree started in the current node
19      def _dfs(v, level):
20          _dfs.h = max(_dfs.h, level)
21          for s in v.sons:
22              if s is not None:
23                  _dfs(s, level + 1)
24      _dfs.h = 0
25      _dfs(self, 0)
26
27      return _dfs.h
28
29  def get_left_path(self):
30      # return size of the leftmost path started in the current node
31      Ln = 1
32      current_vertex = self
33      while len(current_vertex.sons) > 0 \
34          and current_vertex.sons[0] is not None:
35          Ln += 1
36          current_vertex = current_vertex.sons[0]
37      return Ln

```

Лістинг 3.1: Абстрактний клас для дерева.

```

1  # vp_tree.py
2  import tree
3  import sys
4  import math
5
6  sys.setrecursionlimit(20000)
7
8
9  def l2_metric(x, y):
10     ans = 0
11     for i in range(len(x)):
12         ans += (x[i] - y[i])**2
13     return math.sqrt(ans)
14
15
16  def max_metric(x, y):
17     ans = 0
18     for i in range(len(x)):

```

```

19         ans = max(ans, abs(x[i] - y[i]))
20     return ans
21
22
23 def threshold_rule(level):
24     return 1/(level+1)
25
26
27 class VPTree(tree.Tree):
28     # basis VP tree
29
30     def __init__(self, metric=l2_metric, threshold_rule=threshold_rule):
31         super(VPTree, self).__init__()
32         self.sons = [None, None]
33         self.metric = metric
34         self.threshold_rule = threshold_rule
35         self.vantage_point = None
36         self.tau = None
37
38     def add(self, point, level=1):
39         if self.vantage_point is None:
40             self.vantage_point = point
41             self.tau = self.threshold_rule(level)
42         else:
43             if self.metric(point, self.vantage_point) <= self.tau:
44                 i = 0
45             else:
46                 i = 1
47             if self.sons[i] is None:
48                 self.sons[i] = VPTree(metric=self.metric,
49                                     threshold_rule=self.threshold_rule)
50                 self.sons[i].parent = self
51                 self.sons[i].add(point, level=level + 1)
52
53
54 class VPTreeLefts(VPTree):
55     # modified VP tree with h computed as number of left turns
56     # in the path from the root
57
58     def __init__(self, metric=l2_metric, threshold_rule=threshold_rule):
59         super(VPTreeLefts, self).__init__(metric=metric,
60                                     threshold_rule=threshold_rule)
61
62     def add(self, point, lefts=1):
63         if self.vantage_point is None:

```

```

64         self.vantage_point = point
65         self.tau = self.threshold_rule(lefts)
66     else:
67         if self.metric(point, self.vantage_point) <= self.tau:
68             i = 0
69         else:
70             i = 1
71         if self.sons[i] is None:
72             self.sons[i] = VPTreeLefts(metric=self.metric,
73                                       threshold_rule=self.threshold_rule)
74             self.sons[i].parent = self
75             self.sons[i].add(point, lefts=lefts + (1 - i))

```

Лістинг 3.2: Реалізації VP-дерев.

Побудова зображень VP-дерев відбувається додатковою функцією `plot_tree()`, що наведена в лістингу 3.3.

```

1  # vp_tree.py
2
3  # ...
4
5  def plot_tree(plt, tree, N=None, label_nodes=None):
6      # plot a given tree
7      from matplotlib import collections as mc
8      if N is None:
9          n = 1024
10     else:
11         n = N
12     lines = []
13     node_labels = []
14
15     def _dfs(v, level, x, width):
16         p = (x, -level)
17         node_labels.append((x, -level, str(v.vantage_point)))
18         if v.sons[0] is not None:
19             l = (x - width/4, -level-1)
20             lines.append((p, l))
21             _dfs(v.sons[0], level+1, l[0], width/2)
22         if v.sons[1] is not None:
23             r = (x + width/4, -level-1)
24             lines.append((p, r))
25             _dfs(v.sons[1], level+1, r[0], width/2)
26
27     _dfs(tree, 0, 0, n)
28

```

```

29     fig , ax = plt.subplots ()
30     lc = mc.LineCollection (lines , color='b' , linewidths=0.5)
31     ax.add_collection (lc)
32     ax.autoscale ()
33     title = str (type (tree).__name__)
34     if N is not None:
35         title += " , n=%s" % n
36     plt.title (title)
37
38     plt.xticks ([])
39     plt.yticks ([])
40
41     if label_nodes is not None:
42         for label in node_labels:
43             plt.text (label [0]+label_nodes [0] , label [1]+label_nodes [1] ,
44                     label [2] , color="r" , fontsize=14)

```

Лістинг 3.3: Процедура відображення довільного бінарного дерева.

Генерація випадкового дерева та створення його рисунку (див., наприклад, рисунок 3.2) за допомогою цієї функції наведено у лістингу 3.4. Аналогічним чином й відбувалась перевірка гіпотези 58. Для цього достатньо повторити наведену генерацію дерева багато разів та потім усереднити отримані значення `vp_tree.get_height()`.

```

1  import random
2  import matplotlib.pyplot as plt
3  import vp_tree
4
5  n = 1000
6  tau = 0.59
7
8  vp = vp_tree.VPTree (metric=lambda x, y: abs (x - y) ,
9                      threshold_rule=lambda h: tau ** h)
10 for i in range (n):
11     vp.add (random.uniform (-1, 1))
12 vp_tree.plot_tree (plt , vp , n)

```

Лістинг 3.4: Побудова рисунку випадкового VP-дерева.

Дослідження відношення довжини найлівішого шляху та висоти дерева (див. рисунок 3.7) відбувається також з цими класами. У лістингу 3.5 наведено код для моделювання випадкових дерев та підрахунку й відображення

середнього цього відношення.

```
1 import random
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import vp_tree
5
6 TESTS = 50 # number of tests to average values
7 Ns = [1000, 5000, 10000, 20000] # different numbers of nodes to check
8 TAUs = [0.4, 0.5, 0.6, 0.7, 0.8, 0.9] # different params to check
9
10 tree = vp_tree.VPTreeLefts # current tree model
11 for tau in TAUs:
12     E = []
13     for n in Ns:
14         E_Ln_Hn = 0
15         for t in range(TESTS):
16             vp = tree(metric=lambda x, y: abs(x - y),
17                       threshold_rule=lambda h: tau ** h)
18             for i in range(n):
19                 vp.add(2*random.random()-1)
20             Ln = vp.get_left_path()
21             Hn = vp.get_height()
22             E_Ln_Hn += (Ln-1) / Hn
23         E_Ln_Hn /= TESTS
24         E.append(E_Ln_Hn)
25     E = np.array(E)
26     plt.plot(np.log(Ns), E, label="tau=%s" % tau)
27
28 plt.xlabel("log(n)")
29 plt.ylabel("E (Ln-1)/Hn")
30 plt.title(tree.__name__)
31 plt.legend()
```

Лістинг 3.5: Програмний код для генерації випадкових VP-дерев та дослідження відношень довжини найлівішого шляху та висоти.

Окремим кроком дослідження є знаходження функції розподілу в.в. X_∞ . Приклад її статистичного підрахунку й порівняння з теоретичними значеннями наведено у лістингу 3.6.

```
1 import numpy as np
2 from numpy import log
3 import random
4 import matplotlib.pyplot as plt
5
```

```

6
7 def CDF_theory(t, tau):
8     # theoretical value of F(t)
9     if t <= 1.:
10        return 0
11    if t >= 2.:
12        return 1
13    if 1 / tau >= 2.:
14        p = 1. / (1 - 2*log(2) + 1/tau)
15        return p*(t-1)
16    if 1 / tau < 2 and 1 / tau**2 >= 2.:
17        E_1_x = 1/2 / (1 - 4*tau**3 + 4*tau**2 + 2*tau**2*log(2*tau)**2
18                    + 2*tau*log(tau) - 2*tau*log(2*tau))
19        if t <= 1/tau:
20            return 2*tau * (t-1) * E_1_x
21        else:
22            return E_1_x*2*tau*(tau*t-1)
23                    + 2*tau*(t-1)*(E_1_x-2*tau*E_1_x*log(tau*t))
24    return 0
25
26
27 tau = 0.7
28 data = []
29 X = 2 # first point
30 N = 10000 # number of iterations
31 for i in range(N):
32     u = random.uniform(0, 1)
33     X = min(u * X / tau, 1) + min((1 - u) * X / tau, 1)
34     # iteration of the process
35     data.append(X)
36 data.append(3) # adding it for plot purposes at right side
37 plt.hist(data, len(data), histtype='step', cumulative=True,
38         label='Empirical', density=True, linewidth=2)
39
40 x = np.linspace(1, 2.1, 111)
41 plt.plot(x, [CDF_theory(t, tau) for t in x], '.')
42 plt.xlabel('t')
43 plt.ylabel('F(t)')
44 plt.xlim(0.95, 2.05)
45 plt.legend(["Theoretical", "Statistical"], fontsize=20)

```

Лістинг 3.6: Програмний код для емпіричного підрахунку розподілу X_∞ .

3.6 Висновки до розділу 3

В цьому розділі було зроблено перший крок в теоретичному дослідженні структур даних для ефективного пошуку найближчого сусіда, а саме VP-дерев. Попередні відомі в літературі результати щодо цих дерев були отримані суто за допомогою моделювання.

У нашій роботі, зокрема, було отримано наступні результати:

- Досліджено поведінку послідовності розбиттів простору, що породжені найлівішим шляхом випадкового VP-дерева як ланцюга Маркова. Доведено наявність стаціонарного розподілу та збіжність цієї послідовності до нього.
- Отриманий слабкий закон великих чисел для довжини найлівішого шляху VP-дерева.
- Отримана збіжність до випадкової величини для довжини найлівішого шляху вздовж підпослідовностей.
- Отримані точні розподіли для одновимірних розбиттів простору найлівішим шляхом випадкового VP-дерева, коли послідовність точок за якою воно побудоване, має рівномірний розподіл.
- Була висунута гіпотеза щодо балансування випадкового VP-дерева та підтверджена емпіричним шляхом за допомогою моделювання.

Додатково було змодельовано поведінку випадкових VP-дерев й порівняно розподіли одновимірних розбиттів змодельованих дерев та отриманих теоретичних значень.

Оскільки результати цього розділу є лише першим кроком в теоретичному аналізі структур даних для задачі пошуку найближчого сусіда, існує багато напрямків для подальших досліджень. Зокрема, можна виділити наступні:

- Відкритим питанням залишається існування посиленого закону великих чисел для довжини найлівішого шляху випадкового VP-дерева.

- Цікавим напрямком подальших досліджень є знаходження асимптотичного розкладу для довжини найлівішого шляху, в тому числі доведення швидкості збіжності.
- Всі вищенаведені результати варто посилити на випадок довільної метрики в скінченновимірному просторі. Це можливо зробити за допомогою отримання аналогу твердження 37 у випадку довільної метрики, чи, рівносно, іншої форми метричної кулі. Це твердження вже не є таким простим у загальному випадку й зауважимо, що його доведення має скоріш суто геометричний характер. Отримавши це твердження, наступні результати (наприклад, слабкий закон великих чисел) можливо отримати тими самими методами та схемами доведень, що використовувались в цьому розділі.
- Почавши дослідження з розбиттів простору VP-деревом та довжини найлівішого шляху, варто перейти до більш класичних функціоналів, зокрема висоти чи профілю, що дозволять знайти теоретичні оцінки для часу вставки нового елемента та пошуку елемента в дереві. В підрозділі 3.5 емпірично була виявлена залежність між цими величинами та вже дослідженими, тому подальші дослідження можуть складатись з теоретичного знаходження цієї залежності.
- Отримані результати, ідеї та техніки можуть виявитись корисними для вивчення інших типів дерев, що використовуються для розв'язання задачі пошуку найближчого сусіда. Зокрема, в дереві куль виникають подібні до VP-дерева розбиття простору.

ВИСНОВКИ

Дана дисертаційна робота є внеском до теорії випадкових дерев, а саме випадкових дерев простої будови, дерев загальних гіллястих процесів та дерев для задачі пошуку найближчого сусіда. В роботі досліджені асимптотичні властивості цих об'єктів, коли кількість вершин в деревах прямує до нескінченності, й отримано ряд граничних теорем. Отримані теореми як доповнюють вже відомі твердження, посилюючи їх, так і дають принципово нові результати. Використані техніки доведень, допоміжні леми та твердження можна використовувати для подальших досліджень в цьому та суміжних напрямках. Серед отриманих результатів дисертаційної роботи слід виділити наступні:

- За допомогою порівняння з однорозгалудженим гіллястим випадковим блуканням отримано повні асимптотичні розклади майже напевне для профілів випадкових дерев простої будови. Отримано граничні теореми для моди та ширини цих профілів зі збіжністю в сенсі майже напевне. Отримані рівняння нерухомої точки, що характеризують випадкові величини, які фігурують в цих теоремах та розкладах.
- Знайдена асимптотика згортки Лебега-Стільтєса функцій лінійного росту у випадку, коли кількість розглядуваних згортки розбігається до нескінченності. Доведена гранична теорема щодо висоти дерева загального гіллястого процесу, побудованого за збуреним випадковим блуканням.
- Доведено слабкий закон великих чисел для довжини найлівішого шляху випадкового VP-дерева у випадку ℓ_∞ метрики. Також знайдено нетривіальний граничний розподіл цього шляху вздовж підпоследовностей.

Проаналізовано розбиття простору, що породжує VP-дерево, та доведено його граничну збіжність. Сам граничний розподіл додатково охарактеризовано за допомогою рівняння нерухомої точки, що в свою чергу явно розв'язано.

- Наведено комп'ютерне моделювання досліджуваних величин, що дають візуальне представлення про їх еволюцію та підтверджують теоретичні результати.

Додаток

Результати для однорозгалуджених гіллястих випадкових блукань

В цій частині додатку використовуються позначення введені в розділі 1.1. Додамо до них також

$$\mu(\beta) = \varphi'(\beta), \quad \sigma^2(\beta) = \varphi''(\beta)$$

та

$$x_n(k) := x_n(k, \beta) = \frac{k - \mu(\beta) \log n}{\sigma(\beta) \sqrt{\log n}}, \quad k \in \mathbb{Z}.$$

Надалі вважатимемо, що всі 5 тверджень з леми 1 виконуються для однорозгалуджених ГВБ. Наступні твердження запозичені з теорем 3.11, 3.25, 3.17, 3.20 та 3.21 в роботі [43] відповідно, де можна знайти їх доведення.

Теорема 59. *Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) позначає профіль однорозгалудженого ГВБ в момент часу n . Для всіх $r \in \mathbb{N}_0$ та компактної множини $K \subset (\beta_-, \beta_+)$ виконується*

$$(\log n)^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \sup_{\beta \in K} \left| \frac{e^{\beta k}}{n^{\varphi(\beta)}} \mathbb{L}_n(k) - H_n(k; \beta) \sum_{j=0}^r \frac{G_j(x_n(k); \beta)}{(\log n)^{j/2}} \right| \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де G_j наведено в зауваженні 10, та

$$H_n(k; \beta) = \frac{W_\infty(\beta) e^{-\frac{1}{2} x_n^2(k)}}{\sigma(\beta) \sqrt{2\pi \log n}}.$$

Теорема 60. *Нехай $\mathbb{L}_n(k)$ ($k \in \mathbb{Z}$) позначає профіль однорозгалудженого ГВБ з детермінованою кількістю елементів точкового процесу ζ в момент*

часу n . Покладемо $\mathbb{L}_n^\circ(k) := \mathbb{L}_n(k) - \mathbb{E}[\mathbb{L}_n(k)]$ та нехай $(k_n)_{n \in \mathbb{N}}$ позначає детерміновану цілочисельну послідовність.

(а) Якщо $k_n = \mu(0) \log n + \alpha \sigma(0) \log n + o(\sqrt{\log n})$ для деякого $\alpha \in \mathbb{R}$, тоді

$$\frac{\log n}{n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{m(0) \alpha e^{-\frac{1}{2} \alpha^2}}{\sqrt{2\pi \sigma^2(0)}} (\chi_1(0) - \mathbb{E}[\chi_1(0)]).$$

(б) Якщо $k_n = \mu(0) \log n + c_n$, де $c_n = o(\log n)$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді при $\beta_n := \beta_n(k_n)$, що є розв'язком рівняння $\mu(\beta_n) = \frac{k_n}{\log n}$, виконується

$$\frac{(\log n)^{3/2}}{c_n n^{\varphi(\beta_n) - \beta_n \mu(\beta_n)}} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{m(0) (\chi_1(0) - \mathbb{E}[\chi_1(0)])}{\sqrt{2\pi} \sigma^3(0)}.$$

Зокрема, якщо $c_n = o(\sqrt{\log n})$ та $\lim_{n \rightarrow \infty} |c_n| = \infty$, тоді

$$\frac{(\log n)^{3/2}}{n c_n} \mathbb{L}_n^\circ(k_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \frac{m(0) (\chi_1(0) - \mathbb{E}[\chi_1(0)])}{(\sqrt{2\pi} \sigma^3(0))}.$$

(в) Якщо $k_n = \mu(0) \log n + c_n$, де c_n обмежена, тоді

$$\frac{(\log n)^{3/2}}{n} \mathbb{L}_n^\circ(k_n) - R^\circ(c_n) \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 0,$$

де $R^\circ(c) = R(c) - \mathbb{E}R(c)$ та випадкова величина $R(c)$ означена як

$$R(c) := \frac{m(0)}{\sqrt{2\pi} \sigma^3(0)} \left(\chi_1(0) \left(c + \frac{\varphi'''(0)}{2\sigma^2(0)} \right) - \frac{\chi_1^2(0) + \chi_2(0)}{2} \right), \quad c \in \mathbb{R}.$$

Теорема 61. Існує м.н. скінченна випадкова величина K така, що для $n > K$ мода u_n профілю однорозгалудженого ГВБ в момент часу n рівна одному з чисел $\lfloor u_n^* \rfloor$ чи $\lceil u_n^* \rceil$, де

$$u_n^* = \mu(0) \log n + \chi_1(0) - \frac{\varphi'''(0)}{2\sigma^2(0)}.$$

Теорема 62. Нехай M_n позначає ширину профілю однорозгалудженого ГВБ в момент часу n , тоді виконується

$$\frac{\sqrt{2\pi} \log n \sigma(0)}{m(0)n} M_n \xrightarrow[n \rightarrow \infty]{\text{м.н.}} 1.$$

Теорема 63. *Визначимо*

$$\tilde{M}_n := 2\sigma^2(0) \log n \left(1 - \frac{\sqrt{2\pi \log n} \sigma^2(0) M_n}{m(0)n} \right),$$

де M_n має зміст, що й у попередній теоремі. Тоді з $\theta_n = \min_{k \in \mathbb{Z}} |u_n^* - k|$, що фактично означає найближче ціле число до u_n^* , а останнє означено в теоремі 61, виконується

$$\tilde{M}_n - \theta_n^2 \xrightarrow[n \rightarrow \infty]{\text{м.н.}} \chi_2(0) + \frac{(\varphi'''(0))^2}{6\sigma^4(0)} - \frac{\varphi^{(IV)}(0)}{4\sigma^2(0)}.$$

Додаткові твердження

Наступна теорема наведена у повному формулюванні для кращого розуміння схеми доведення леми 51.

Теорема 64. *Нехай дано два набори випадкових величин $(X_{m,n})_{m \in \mathbb{N}, n \in \mathbb{N}}$ та $(Y_n)_{n \in \mathbb{N}}$ такі, що для кожного $n \in \mathbb{N}$ елементи $Y_n, X_{1,n}, X_{2,n}, \dots$ мають спільну область визначення, а їх простір значень сепарабельний. Нехай виконується $X_{m,n} \xrightarrow{d} X_m$ при $n \rightarrow \infty$ та $X_m \xrightarrow{d} X$ при $m \rightarrow \infty$, а також для будь-якого $\varepsilon > 0$*

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\{\rho(X_{m,n}, Y_n) \geq \varepsilon\} = 0,$$

де $\rho(\cdot, \cdot)$ позначає відстань у відповідному просторі значень.

Тоді $Y_n \xrightarrow{d} X$ при $n \rightarrow \infty$.

Доведення. Доведення цієї теореми можна знайти в книзі [9], див. теорему 3.2 на с. 28. □

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] G. Ambrus, P. Kevei, and V. Vigh. The diminishing segment process. *Statist. Probab. Lett.*, 82(1):191–195, 2012.
- [2] S. Asmussen. *Applied probability and queues, Second edition*. Springer-Verlag, 2003.
- [3] F. Ball, M. Gonzalez, R. Martinez, and M. Slavtchova-Bojkov. Stochastic monotonicity and continuity properties of functions defined on Crump-Mode-Jagers branching processes, with application to vaccination in epidemic modelling. *Bernoulli*, 20(4):2076–2101, 2014.
- [4] R.E. Barlow and F. Proscha. *Statistical Theory of Reliability and Life Testing: Probability Models. To Begin With*, 1981.
- [5] F. Bavaud. Adjoint transform, overconvexity and sets of constant width. *Trans. Amer. Math. Soc.*, 333(1):315–324, 1992.
- [6] J. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [7] S. Bhamidi, R. Hofstad, and J. Komjathy. The front of the epidemic spread and first passage percolation. *J. Appl. Probab.*, 51A(Celebrating 50 Years of The Applied Probability Trust):101–121, 2014.
- [8] J. D. Biggins. Chernoff’s theorem in the branching random walk. *J. Appl. Probab.*, 14:630–636, 1977.
- [9] P. Billingsley. *Convergence of Probability Measures, Second Edition*. John Wiley & Sons, New York, 1999.

- [10] V. Bogun. Almost sure asymptotic expansions for profiles of simply generated random trees. *Theory of Stochastic Processes*, 24(40)(1):49–63, 2019.
- [11] V. Bohun. Probabilistic analysis of vantage point trees. *Modern Stochastics: Theory and Applications*, 8(4):413–434, 2021.
- [12] V. Bohun, A. Iksanov, A. Marynych, and B. Rashytov. Renewal theory for iterated perturbed random walks on a general branching process tree: intermediate generations. *Journal of Applied Probability*, 59(2):421–446, 2022.
- [13] L. Boytsov and B. Naidan. Learning to prune in metric and non-metric spaces. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 1:1574–1582, 2013.
- [14] D. Buraczewski, B. Dovgay, and A. Iksanov. On intermediate levels of nested occupancy scheme in random environment generated by stick-breaking i. *Electron. J. Probab.*, 25:1–24, 2020.
- [15] B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *Ann. Appl. Probab.*, 11(4):1042–1062, 2001.
- [16] B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault. Martingales and profile of binary search trees. *Elect. J. Probab.*, 10:420–435, 2005.
- [17] L. Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. Reidel Publishing Company, 1974.
- [18] T. Cover. Estimation by the nearest neighbor rule. *IEEE Transactions on Information Theory*, 14(1):50–55, 1968.
- [19] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.
- [20] L. Devroye. The uniform convergence of nearest neighbor regression function estimators and their applications in optimization. *IEEE Transactions on Information Theory*, 24(2):142–151, 1978.

- [21] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.
- [22] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Inform.*, 24(3):277–298, 1987.
- [23] L. Devroye and H.-K. Hwang. Width and mode of the profile for some random trees of logarithmic height. *Ann. Appl. Probab.*, 16(2):886–918, 2006.
- [24] M. Drmota. *Random trees: An interplay between combinatorics and probability*. Springer–Verlag Wien, 2009.
- [25] M. Drmota and H.-K. Hwang. Profiles of random trees: Correlation and width of random recursive trees and binary search trees. *Advances in Applied Probability*, 37(2):321–341, 2005.
- [26] M. Drmota, S. Janson, and R. Neininger. A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.*, 18(1):288–333, 2008.
- [27] T. Duquesne and J.-F. Le Gall. *Random trees, Lévy processes and spatial branching processes*. Société mathématique de France, 2002.
- [28] R. Durrett. *Probability Theory and Examples, Fourth Edition*. Cambridge University Press, New York, 2010.
- [29] V. Feray, P.-L. Meliot, and A. Nikeghbali. *Mod- ϕ Convergence: Normality Zones and Precise Deviations*. SpringerBriefs in Probability and Mathematical Statistics. Springer International Publishing, 2016.
- [30] J. B. G. Frenk. *On Banach algebras, renewal measures and regenerative processes*. Stichting Mathematisch Centrum, 1987.
- [31] A. Fu, P. Chan, Y. Cheung, and Y. Moon. Dynamic vp-tree indexing for n-nearest neighbor search given pair-wise distances. *The VLDB Journal*, 9:154–173, 2000.

- [32] M. Fuchs, H.-K. Hwang, and R. Neininger. Profiles of random trees: limit theorems for random recursive trees and binary search trees. *Algorithmica*, 46(3-4):367–407, 2006.
- [33] R. Gouet. Strong convergence of proportions in a multicolor Pólya urn. *Journal of Applied Probability*, 34(2):426–435, 1997.
- [34] C. Holmgren and S. Janso. Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees. *Probab. Surv.*, 14:53–154, 2017.
- [35] H.-K. Hwang. Profiles of random trees: plane-oriented recursive trees. *Random Structures Algorithms*, 30(3):380–413, 2007.
- [36] A. Iksanov and Z. Kabluchko. A functional limit theorem for the profile of random recursive trees. *Electronic Communications in Probability*, 23(87):1–13, 2018.
- [37] A. Iksanov and Z. Kabluchko. Weak convergence of the number of vertices at intermediate levels of random recursive trees. *Journal of Applied Probability*, 55(4):1131–1142, 2018.
- [38] A. Iksanov and M. Möhle. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Probab.*, 12:28–35, 2007.
- [39] J. Jacod, E. Kowalski, and A. Nikeghbali. Mod-gaussian convergence: new limit theorems in probability and number theory. *Forum Math.*, 23(4):835–873, 2011.
- [40] P. Jagers. *Branching processes with biological applications*. John Wiley & Sons, 1975.
- [41] P. Jagers and S. Sagitov. General branching processes in discrete time as random trees. *Bernoulli*, 14(4):949–962, 2008.

- [42] V. Jog and P.-L. Loh. Analysis of centrality in sublinear preferential attachment trees via the Crump-Mode-Jagers branching process. *EEE Trans. Network Sci. Eng.*, 4(1):1–12, 2017.
- [43] Z. Kabluchko, A. Marynych, and H. Sulzbach. General edgeworth expansions with applications to profiles of random trees. *The Annals of Applied Probability*, 27(6):3478–3524, 2017.
- [44] Z. Katona. Width of a scale-free tree. *J. Appl. Probab.*, 42(3):839–850, 2005.
- [45] P. Kevei and V. Vigh. On the diminishing process of Balint Toth. *Trans. Amer. Math. Soc.*, 368(12):8823–8848, 2016.
- [46] M. Kimmel and D. E. Axelrod. *Branching processes in biology, volume 19 of Interdisciplinary Applied Mathematics, 2nd edition*. Springer, 2015.
- [47] D. E. Knuth. *The Art of Computer Programming. Vol. 3: Sorting and Searching*. Addison-Wesley Publishing Company, 1998.
- [48] E. Kowalski and A. Nikeghbali. Mod-poisson convergence in probability and number theory. *Int. Math. Res. Not. IMRN*, 18:3549–3587, 2010.
- [49] N. Kumar, L. Zhang, and S. Nayar. What is a good nearest neighbors algorithm for finding similar patches in images. *Lecture Notes in Computer Science*, 5303:364–378, 2008.
- [50] K. Leckey, D. Mitsche, and N. Wormald. The height of depth-weighted random recursive trees. *Random Structures Algorithms*, 56(3):851–866, 2020.
- [51] S. S. Lee, M. Shishibori, and C. Y. Han. Compression of search range of vp-tree for multimedia data retrieval applications. *2012 Data Compression Conference*, page 403, 2012.
- [52] W. Li, Y. Zhang, Y. Sun, W. Wang, W. Zhang, and X. Lin. Approximate nearest neighbor search on high dimensional data — experiments, analyses, and improvement. *IEEE Transactions on Knowledge and Data Engineering*, 32:1475–1488, 2020.

- [53] S. Maghsoodloo and D. Helvaci. Renewal and renewal-intensity functions with minimal repair. *Journal of Quality and Reliability Engineering*, 2014.
- [54] H. Mahmoud. Profile of random exponential recursive trees. *Methodology and Computing in Applied Probability*, pages 1–17, 2021.
- [55] H. Mahmoud. *Pólya urn models*. CRC Press, 2009.
- [56] I. Markov. Vp-tree: Content-based image indexing. *SYRCoDIS*, 2007.
- [57] C. Martinez and S. Roura. Randomized binary search trees. *Journal of the ACM*, 45(2):288–323, 1998.
- [58] A. Meir and J. W. Moon. Cutting down random trees. *Journal of the Australian Mathematical Society*, 11(3):313–324, 1970.
- [59] A. Moore. An introductory tutorial on kd-trees. *Technical Report No. 209, Computer Laboratory, University of Cambridge*, 1991.
- [60] Y. Narasimhulu, A. Suthar, R. Pasunuri, and V. Venkaiah. Ckd-tree: An improved kd-tree construction algorithm. *International Semantic Intelligence Conference 2021*, pages 211–218, 2021.
- [61] Frank Nielsen, Paolo Piro, and Michel Barlaud. Bregman vantage point trees for efficient nearest neighbor queries. *2009 IEEE International Conference on Multimedia and Expo*, pages 878–881, 2009.
- [62] S. Omohundro. Five balltree construction algorithms. *Technical Report 89-063*, 1989.
- [63] N. Papernot and P. McDaniel. Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. *CoRR*, abs/1803.04765, 2018.
- [64] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.
- [65] B. Reed. The height of a random binary search tree. *Journal of the ACM*, 50(3):306–332, 2003.

- [66] U. Roster. A limit theorem for "quicksort". *RAIRO Inform. Theor. Appl.*, 25(1):85–100, 1991.
- [67] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures and Algorithms*, 31(2):186–202, 2007.
- [68] W. Rudin. *Fourier analysis on groups*. John Wiley & Sons, 1962.
- [69] L. Sasongko and T. Mahatma. The estimation of renewal functions using the mean value theorem for integrals (MeVTI) method. *d’CARTESIAN*, 5(2):111–120, 2016.
- [70] E.-M. Schopp. A functional limit theorem for the profile of b-ary trees. *Ann. Appl. Probab.*, 20(3):907–950, 2010.
- [71] P. K. Sen. Weak convergence of an iterated renewal process. *J. Appl. Probab.*, 18:291–296, 1981.
- [72] M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer Science+Business Media, New York, 2007.
- [73] M. Shishibori, S. Lee, and K. Kita. An improved method to select candidates on metric index vp-tree. *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, 1:306–311, 2011.
- [74] M. Shishibori, S. S. Lee, and K. Kita. A method to improve metric index vp-tree for multimedia databases. *17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, pages 1–6, 2011.
- [75] R. Smythe and H. Mahmoud. A survey of recursive trees. *Theory Probab. Math. Statist.*, 51:1–27, 1995.
- [76] H. Sulzbach. A functional limit law for the profile of plane-oriented recursive trees. *In Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc., AI, Assoc. Discrete Math. Theor. Comput. Sci.*, pages 339–350, 2008.

- [77] J. Vermorel. Near neighbor search in metric and nonmetric space. *hal-00004887v1*, 2005.
- [78] M. Xie. On the solution of renewal-type integral equations. *Communication Statistics Simulation Computation*, 18(1):281–293, 1989.
- [79] P. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. *Fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 311–321, 1993.

Онлайн сервіс створення та перевірки кваліфікованого та удосконаленого електронного підпису

ПРОТОКОЛ
створення та перевірки кваліфікованого та удосконаленого електронного підпису

Дата та час: 15:47:33 09.08.2022

Назва файлу з підписом: bohun_phd.pdf.asice
Розмір файлу з підписом: 1.4 МБ

Перевірені файли:
Назва файлу без підпису: bohun_phd.pdf
Розмір файлу без підпису: 1.5 МБ

Результат перевірки підпису: Підпис створено та перевірено успішно. Цілісність даних підтверджено

Підписувач: БОГУН ВЛАДИСЛАВ АНАТОЛІЙОВИЧ
П.І.Б.: БОГУН ВЛАДИСЛАВ АНАТОЛІЙОВИЧ
Країна: Україна
РНОКПП: 3451000936
Організація (установа): ФІЗИЧНА ОСОБА
Час підпису (підтверджено кваліфікованою позначкою часу для підпису від Надавача): 15:47:31 09.08.2022
Сертифікат виданий: АЦСК АТ КБ «ПРИВАТБАНК»
Серійний номер: 2B6C7DF9A3891DA1040000001CEBCE003DFA9003
Алгоритм підпису: ДСТУ-4145
Тип підпису: Удосконалений
Тип контейнера: Підпис та дані в архіві (розширений) (ASiC-E)
Формат підпису: З повними даними для перевірки (XAdES-B-LT)
Сертифікат: Кваліфікований