

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

**Фізико-математичний факультет**

**Кафедра математичного аналізу та теорії ймовірностей**

«На правах рукопису»

УДК 519.21

До захисту допущено  
завідувач кафедри

О.І. Клесов

« \_\_\_\_ » \_\_\_\_\_ 2024 р.

**Магістерська дисертація**

**на здобуття ступеня магістра**

**за освітньо-науковою програмою «Страхова та фінансова математика»**

**зі спеціальності 111 «Математика»**

**на тему: «Граничні теореми для випадкових перестановок з вагами циклів»**

Виконав:

студент 2 курсу магістратури, групи ОМ-21мн

Галганов Олексій Андрійович

\_\_\_\_\_

Науковий керівник:

доцент кафедри МАтаТЙ ФМФ КПІ,

кандидат фізико-математичних наук

Ільєнко Андрій Борисович

\_\_\_\_\_

Рецензент:

доцент кафедри ММСА ІПСА КПІ,

кандидат фізико-математичних наук

Каніовська Ірина Юріївна

\_\_\_\_\_

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів без  
відповідних посилань.

Студент: \_\_\_\_\_

Київ – 2024 року

**Національний технічний університет України**  
**«Київський політехнічний інститут імені Ігоря Сікорського»**  
**Фізико-математичний факультет**  
**Кафедра математичного аналізу та теорії ймовірностей**

Рівень вищої освіти – другий (магістерський)

Спеціальність – 111 «Математика»

Освітньо-наукова програма «Страхова та фінансова математика»

ЗАТВЕРДЖУЮ  
завідувач кафедри  
\_\_\_\_\_ О.І. Клесов  
« \_\_\_\_ » \_\_\_\_\_ 2024 р.

**ЗАВДАННЯ**  
**на магістерську дисертацію студенту**  
**Галганову Олексію Андрійовичу**

1. Тема дисертації «Граничні теореми для випадкових перестановок з вагами циклів», науковий керівник дисертації Ільєнко Андрій Борисович, доцент, кандидат фізико-математичних наук, затверджені наказом по університету від 1 квітня 2024 р., №1523-с.
2. Термін подання студентом дисертації: 11 червня 2024 року.
3. Об'єкт дослідження: випадкові перестановки з вагами циклів.
4. Предмет дослідження: граничні теореми для точкових процесів, породжених циклами випадкових перестановок.
5. Перелік завдань, які потрібно розробити:
  - 1) Ознайомитись з відомими результатами, що стосуються граничних теорем для статистик випадкових перестановок з вагами циклів.
  - 2) Побудувати точковий випадковий процес, який описує цикли випадкових перестановок.
  - 3) Довести граничну теорему для побудованої послідовності точкових процесів, попередньо встановивши, за яких умов на параметри розподілу це має сенс.
  - 4) Ознайомитись з теоремою про неперервне відображення для точкових випадкових процесів.
  - 5) Сформулювати та довести граничні теореми для деяких статистик випадкових перестановок.

- б) Ознайомитись з методами чисельного моделювання, провести відповідні симуляції для перевірки отриманих результатів.
6. Орієнтовний перелік графічного (ілюстративного) матеріалу: 28 слайдів.
7. Перелік публікацій:
- 1) Тези доповіді на XI Всеукраїнській науковій конференції молодих математиків, Київ, 11-13 травня 2023 р.
  - 2) Тези доповіді на XIX Міжнародній науковій конференції імені академіка Михайла Кравчука, Київ, 11-12 жовтня 2023 р.
  - 3) Тези доповіді на XII Всеукраїнській науковій конференції молодих математиків, Київ, 9-11 травня 2024 р.
  - 4) Galganov O., Plienko A. Short cycles of random permutations with cycle weights: Point processes approach. // *Statistics & Probability Letters*. 2024. Vol. 213. DOI: <https://doi.org/10.1016/j.spl.2024.110169>.
8. Дата видачі завдання: 5 лютого 2024 року.

#### Календарний план

№ з/п	Назва етапу виконання магістерської дисертації	Термін виконання етапу магістерської дисертації	Примітка
1	Огляд літератури за темою	до 04.03.2024	виконано
2	Побудова точкового процесу	до 11.03.2024	виконано
3	Доведення граничної теореми для послідовності точкових процесів	до 08.04.2024	виконано
4	Формулювання та доведення граничних теорем для статистик	до 06.05.2024	виконано
5	Чисельне моделювання	до 13.05.2024	виконано
6	Оформлення магістерської дисертації	до 03.06.2024	виконано

Студент  
Науковий керівник

Олексій ГАЛГАНОВ  
Андрій ІЛЬЄНКО

## РЕФЕРАТ

Магістерська дисертація містить 37 сторінок, 11 рисунків, 20 джерел, 1 додаток, 28 слайдів презентації.

Останніми роками після десятиліть певного забуття в теорії ймовірностей відроджується інтерес до тематики точкових випадкових мір та їхніх застосувань. Точкові випадкові міри, також відомі як точкові процеси, викликають інтерес не тільки потужною й красивою математичною теорією, а й новими можливостями для дослідження дискретних моделей з класичної теорії ймовірностей.

Об'єктом дослідження в роботі є так звані випадкові перестановки з вагами циклів, які останнім часом здобули певну популярність в літературі, особливо в контексті застосувань до задач статистичної фізики. Також, одним з варіантів цієї моделі є перестановки Юенса, які було вперше досліджено в 70-их роках ХХ ст. в роботах з генетики популяцій.

Метою роботи є отримання та доведення граничної теореми для послідовності точкових процесів, породжених циклами випадкових перестановок, а також – граничних теорем для деяких статистик циклів.

Дослідження передбачає роботу з науковою літературою за темою, зокрема використання теоретичної бази теорії точкових випадкових мір та відомих результатів, що стосуються циклів випадкових перестановок.

В даній роботі пропонується новий підхід для дослідження вищезгаданих випадкових перестановок, а саме – аналіз асимптотики певного точкового процесу, пов'язаного з циклами перестановок. Фактично, пропонується досліджувати певний «геометричний» опис перестановок, а не лише «арифметичний», як у відомих роботах.

Результати з магістерської дисертації були представлені на конференціях: XI Всеукраїнській науковій конференції молодих математиків (Київ, травень 2023 р.), XIX Міжнародній науковій конференції імені академіка Михайла Кравчука (Київ, жовтень 2023 р.), XII Всеукраїнській науковій конференції молодих математиків (Київ, травень 2024 р.). Робота є переможцем I туру Всеукраїнського конкурсу студентських робіт з галузей знань і спеціальностей у 2023/2024 навчальному році. Статтю, що висвітлює основні результати роботи, прийнято до публікації в журналі *Statistics & Probability Letters*.

Ключові слова: *випадкові перестановки, перестановки Юенса, точкові процеси, процес Пуассона, груба збіжність за розподілом.*

## ABSTRACT

Master's thesis: 37 pages, 11 figures, 20 references, 1 appendix, 28 presentation slides.

After decades of being somewhat forgotten, random point measures (also known as point processes) and their applications are reviving in popularity among the scientific community. Not only the powerful and beautiful theory but also new possibilities for studying classical discrete probability models explain the interest in this topic.

The object of the research is the model of random permutations with cycle weights, which is gaining popularity in literature, especially in applications to statistical physics. Also, Ewens permutations, first studied in the 1970s in the context of population genetics, are a special case of this model.

The goal of this work is to obtain and prove a limit theorem for a sequence of point processes generated by the cycles of a random permutation, together with limit theorems for a few cycles' statistics.

Research methods are based on extensive use of scientific literature, including basic theory of random point measures, and known results regarding cycles of random permutations.

This work introduces a novel approach for studying the aforementioned model, namely the study of the asymptotic behaviour of some point process directly related to cycles of a permutation. In other words, it focuses on a "geometric" representation of a permutation rather than an "arithmetic" one suggested by known works.

The results of this work were presented at the following events: XI All-Ukrainian Scientific Conference of Young Mathematicians (Kyiv, May 2023), XIX International Scientific Mykhailo Kravchuk Conference (Kyiv, October 2023), XII All-Ukrainian Scientific Conference of Young Mathematicians (Kyiv, May 2024). This work is the winner of the 1st round of the All-Ukrainian contest of student works in the fields of knowledge of the 2023-2024 academic year. A paper has been accepted for publication in the *Statistics & Probability Letters*.

Keywords: *random permutations, Ewens permutations, point processes, Poisson process, vague convergence in distribution.*

## ЗМІСТ

СКРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ.....	7
ВСТУП .....	8
РОЗДІЛ 1 Огляд відомих результатів .....	10
РОЗДІЛ 2 Основна теорема та її наслідки.....	12
2.1. Деякі теоретичні відомості .....	12
2.2. Основна теорема .....	14
2.3. Наслідки та приклади застосування .....	18
РОЗДІЛ 3 Чисельне моделювання отриманих результатів .....	23
3.1. Процес китайського ресторану.....	23
3.2. Моделювання результатів.....	24
ВИСНОВКИ .....	29
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	30
ДОДАТОК А. ПРОГРАМНИЙ КОД ДЛЯ МОДЕЛЮВАННЯ .....	32

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ

$\#A$	потужність множини $A$
$[n]$	множина $\{1, \dots, n\}$
$\mathcal{S}_n$	група перестановок (симетрична група) степеня $n$
$C_k(\pi)$	кількість циклів довжини $k$ у розкладі перестановки $\pi$ в композицію незалежних циклів
$\mathbb{1}\{\cdot\}$	індикаторна функція, що дорівнює 1 у випадку, коли умова в дужках справджується, і 0 у іншому випадку
$\mathcal{B}(X)$	борелева $\sigma$ -алгебра на множині $X$
$\widehat{\mathcal{B}}(X)$	сім'я обмежених борелевих підмножин $X$
$\langle a, b \rangle$	інтервал, позначає одне з $[a, b]$ , $(a, b)$ , $[a, b)$ , $(a, b]$
$\lambda_k$	міра Лебега в $\mathbb{R}^k$ або її звуження на довільну $A \in \mathcal{B}(\mathbb{R}^k)$
$\delta_x$	міра Дірака, зосереджена в точці $x$
$M_p(X)$	множина локально скінченних точкових мір на $(X, \mathcal{B}(X))$
$a_n \rightarrow a$	числова послідовність $a_n$ збігається до $a$
$\mu_n \xrightarrow{v} \mu$	послідовність мір $\mu_n$ грубо збігається до міри $\mu$
$\xi_n \xrightarrow{vd} \xi$	послідовність точкових випадкових процесів $\xi_n$ грубо збігається за розподілом до точкового випадкового процесу $\xi$
$X_n \xrightarrow{d} X$	послідовність випадкових величин $X_n$ збігається за розподілом до випадкової величини $X$
$X \stackrel{d}{=} Y$	випадкові величини $X$ та $Y$ рівні за розподілом
$X_{(k)}$	$k$ -та порядкова статистика, тобто $k$ -та за номером випадкова величина серед відсортованих у порядку зростання неперервних випадкових величин $X_1, X_2, \dots, X_n$ , $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$
$\mathbb{E}X$	математичне сподівання випадкової величини $X$
$X \sim D$	випадкова величина $X$ має розподіл $D$
$\text{Pois}(a)$	розподіл Пуассона з параметром $a > 0$ , $\mathbb{P}(X = n) = \frac{a^n}{n!} e^{-a}$ для $n \in \mathbb{N} \cup \{0\}$
$\text{Exp}(\lambda)$	експоненційний розподіл з параметром $\lambda > 0$ зі щільністю $f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}\{x \geq 0\}$
$\text{ESF}(n, \theta)$	розподіл Юенса з параметром $\theta$ на $\mathcal{S}_n$
$I_\nu(z)$	модифікована функція Бесселя першого роду, $\nu \in \mathbb{R}$

## ВСТУП

Основним об'єктом досліджень в цій роботі є випадкові перестановки. *Перестановка*  $\sigma$  – це взаємно-однозначне перетворення набору натуральних чисел  $[n] = \{1, \dots, n\}$ , яке часто записують, перелічуючи значення на кожному  $i \in [n]$ :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix},$$

або ж просто  $\sigma = (\sigma(1), \sigma(2), \sigma(3), \dots, \sigma(n))$ . Наприклад,  $\sigma = (2, 1, 4, 3, 5)$  означає, що відображення  $\sigma : [5] \rightarrow [5]$  діє за правилом  $\sigma(1) = 2, \sigma(2) = 1, \sigma(3) = 4, \sigma(4) = 3, \sigma(5) = 5$ . Множину всіх перестановок на  $[n]$  позначатимемо  $\mathcal{S}_n$ .

З-поміж багатьох математичних властивостей перестановок, з якими можна ознайомитись, наприклад, в [20], окремо варто звернути увагу на можливість розкладу будь-якої перестановки в *композицію незалежних циклів* (напр., теорема 6.3 в [20]). Це означає, що для довільної перестановки  $\sigma \in \mathcal{S}_n$  існує таке розбиття  $[n]$  на підмножини  $\{i_{1,1}, \dots, i_{1,k_1}\}, \{i_{2,1}, \dots, i_{2,k_2}\}, \dots, \{i_{m,1}, \dots, i_{m,k_m}\}$ , що не перетинаються і в об'єднанні дають  $[n]$ , на кожній з яких дія перестановки спрощується до  $\sigma(i_{j,1}) = i_{j,2}, \sigma(i_{j,2}) = i_{j,3}, \dots, \sigma(i_{j,k_j-1}) = i_{j,k_j}, \sigma(i_{j,k_j}) = i_{j,1}$ . Це дає змогу записувати кожен перестановку як

$$\sigma = (i_{1,1}, \dots, i_{1,k_1}) \circ (i_{2,1}, \dots, i_{2,k_2}) \circ \dots \circ (i_{m,1}, \dots, i_{m,k_m}),$$

таким записом ми будемо часто користуватись надалі. Числа  $k_j$  називають довжинами циклів. Наприклад, для згаданої вище перестановки  $\sigma = (2, 1, 4, 3, 5)$  маємо  $\sigma = (1, 2) \circ (3, 4) \circ (5)$  – можна сказати, що ця перестановка має один цикл довжини 1 та два цикли довжини 2.

В роботі розглядаються випадкові перестановки  $\sigma_n \in \mathcal{S}_n$ , задані розподілом

$$\mathbb{P}(\sigma_n = \sigma) = \frac{1}{h_n n!} \prod_{k=1}^{\infty} \theta_k^{C_k(\sigma)}, \quad \sigma \in \mathcal{S}_n, \quad (1)$$

де  $\theta_k$  – невід'ємні дійсні параметри,  $C_k(\sigma)$  – кількість циклів довжини  $k$  в розкладі перестановки  $\sigma$  ( $C_k(\sigma) = 0$  при  $k > n$ ), а  $h_n$  – така стала, що забезпечує  $\sum_{\sigma \in \mathcal{S}_n} \mathbb{P}(\sigma_n = \sigma) = 1$ . Зауважимо, що іноді в літературі (напр., [3]) використовується еквівалентне означення

$$\mathbb{P}(\sigma_n = \sigma) = \frac{1}{h_n n!} \exp \left\{ \sum_{k=1}^{\infty} \alpha_k C_k(\sigma) \right\}, \quad \sigma \in \mathcal{S}_n.$$



Формула (1) є найбільш загальною формою запису розподілу так званих *випадкових перестановок з вагами циклів*. Найпростішими прикладами є рівномірний розподіл при  $\theta_k \equiv 1$  та розподіл Юенса при  $\theta_k \equiv \theta > 0$  ([15, 19]). В усіх інших випадках нас цікавитиме лише асимптотична поведінка сталих  $h_n$  при  $n \rightarrow \infty$ .

Варто зауважити, що параметри розподілу (1) насправді визначені з точністю до множників виду  $e^{ak}$  для  $a \in \mathbb{R}$ . Дійсно, якщо зробити заміну  $\theta_k \mapsto e^{ak}\theta_k$ , то

$$\prod_{k=1}^{\infty} (e^{ak}\theta_k)^{C_k(\sigma)} = \exp \left\{ a \sum_{k=1}^{\infty} kC_k(\sigma) \right\} \prod_{k=1}^{\infty} \theta_k^{C_k(\sigma)} = e^{an} \prod_{k=1}^{\infty} \theta_k^{C_k(\sigma)},$$

оскільки для будь-якої перестановки  $\sigma \in \mathcal{S}_n$  виконується  $\sum_{k=1}^n kC_k(\sigma) = n$ .

Відомі результати, які детально буде розглянуто далі, стосуються різних числових характеристик циклів, що входять до складу випадкової перестановки, як от кількості циклів заданої довжини чи довжини типового циклу. Метою нашої роботи є дослідження «геометрії» циклів при  $n \rightarrow \infty$ . Інакше кажучи, головне питання звучить так: *як описати склад циклів випадкової перестановки великої довжини?*

Надзвичайно цікавим є те, що формула (1) (звісно, для різних наборів  $\theta_k$ ) застосовується для опису математичних моделей у багатьох дисциплінах. Так, розподіл Юенса вперше виник в математичних дослідженнях з генетики популяцій ([19, 17, 6]), а розподіли з певною асимптотикою  $\theta_k$  застосовуються в моделях квантової механіки для опису конденсації Бозе-Ейнштейна ([1, 3, 2]). Усвідомлюючи свою недостатню обізнаність з теоретичною базою інших дисциплін, в яких отримані результати можуть мати застосування, ми утримаємось від детальних пояснень, обмежившись наведеними посиланнями на окремі роботи, які можуть бути корисними для спеціалістів.

## РОЗДІЛ 1

### ОГЛЯД ВІДОМИХ РЕЗУЛЬТАТІВ

Випадкові перестановки є класичним об'єктом комбінаторної теорії ймовірностей. Історія досліджень починається принаймні з XVIII століття. У 1708 році П'єр де Монмор у праці «Essay d'analyse sur les jeux de hazard» («Аналітичний нарис про азартні ігри», [12]) отримав формулу для ймовірності того, що принаймні одна з нумерованих карт з колоди з  $n$  штук при випадковому тасуванні опиниться на своєму місці:

$$1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1} \frac{1}{n!}.$$

Нескладно побачити, що при  $n \rightarrow \infty$  ця ймовірність прямує до  $1 - e^{-1}$ . За допомогою простих комбінаторних міркувань можна також отримати формулу для ймовірності того, що на своїх місцях залишиться рівно  $m$  карт з  $n$ :

$$\frac{1}{m!} \left( 1 - \left( 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-m-1} \frac{1}{(n-m)!} \right) \right) = \frac{1}{m!} \sum_{j=0}^{n-m} \frac{(-1)^j}{j!}.$$

Тут при  $n \rightarrow \infty$  границя теж існує, і дорівнює  $\frac{1}{m!} e^{-1}$ .

Цей результат можна сформулювати так: якщо випадкова перестановка  $\sigma_n$  має розподіл (1) з  $\theta_k \equiv 1$ , то  $\lim_{n \rightarrow \infty} \mathbb{P}(C_1(\sigma_n) = m) = \mathbb{P}(X = m)$ , де  $X$  – випадкова величина, що має розподіл Пуассона з параметром 1,  $X \sim \text{Pois}(1)$ . Інакше кажучи, кількість циклів довжини 1 у рівномірній випадковій перестановці  $n$  елементів при  $n \rightarrow \infty$  збігається за розподілом до  $\text{Pois}(1)$ .

Першим логічним узагальненням цього результату є відоме твердження про сумісний розподіл кількості циклів всіх можливих довжин у рівномірній випадковій перестановці  $\sigma_n$  (напр., теорема 1.3 в [15]):

$$(C_1(\sigma_n), C_2(\sigma_n), C_3(\sigma_n), \dots) \xrightarrow{d} (X_1, X_2, X_3, \dots), \quad n \rightarrow \infty, \quad (2)$$

де  $X_k \sim \text{Pois}(1/k)$  є незалежними випадковими величинами, а  $\xrightarrow{d}$  означає збіжність за розподілом в  $\mathbb{Z}_+^\infty$  з топологією поточної збіжності. Зауважимо, що мова йде лише про цикли фіксованої довжини, які іноді називають *короткими циклами*. Дослідження величини типу  $C_{\varphi(n)}(\sigma_n)$ , де довжина циклу залежить від  $n$  (так звані *довгі цикли*), є зовсім іншою задачею.

Доволі відомим фактом є аналог (2) для вже згаданих перестановок Юенса з  $\theta_k \equiv \theta > 0$  ([14], теорема 1), де  $X_k \sim \text{Pois}(\theta/k)$  і теж є незалежними.

Наступним кроком узагальнення (2) є випадок довільних  $\theta_k$ . В [8] доведено, що достатньою умовою в цьому разі є  $\lim_{n \rightarrow \infty} \frac{h_{n-1}}{h_n} = 1$ , і також показано, що ця умова виконується для доволі широкого класу асимптотик параметрів  $\theta_k$  – від субекспоненційного спадання до субекспоненційного зростання. В цьому випадку  $X_k$  теж будуть незалежними та матимуть розподіл  $\text{Pois}(\theta_k/k)$ .

Як вже було сказано, згадані вище результати описують лише арифметичні властивості циклів – їхню кількість чи довжину. При цьому для формального опису використовується відносно простий та, безперечно, відомий апарат теорії ймовірностей – випадкові величини. Для досягнення основної мети роботи, опису геометрії коротких циклів, ми будемо використовувати складніші об'єкти – *точкові випадкові процеси*.

## РОЗДІЛ 2

## ОСНОВНА ТЕОРЕМА ТА ЇЇ НАСЛІДКИ

## 2.1. Деякі теоретичні відомості

Як вже було зазначено, перестановку можна повністю описати складом елементів, які входять в її незалежні цикли. Для дослідження випадкових перестановок необхідно формалізувати поняття «випадковий набір точок».

Детальніше з відповідною теорією можна ознайомитись, наприклад, в [16] чи [10]. Ми ж наведемо лише деякі означення та пояснення, яких має бути достатньо, щоб сформулювати уявлення про об'єкти, які будуть досліджуватися.

Нехай  $\mathbb{X}$  – повний сепарабельний метричний простір,  $\mathcal{B}(\mathbb{X})$  –  $\sigma$ -алгебра борелевих підмножин,  $\widehat{\mathcal{B}}(\mathbb{X})$  – сім'я обмежених множин з  $\mathcal{B}(\mathbb{X})$ ,  $M_p(\mathbb{X})$  – множина локально скінченних точкових мір на  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ ,  $\mathcal{M}_p(\mathbb{X})$  – найменша  $\sigma$ -алгебра, що робить вимірними всі відображення виду  $\varphi_B(\mu) = \mu(B)$  для  $B \in \mathcal{B}(\mathbb{X})$ , а  $(\Omega, \mathcal{A}, \mathbb{P})$  – деякий простір елементарних подій.

**Означення 1.** *Точковий процес (точкова випадкова міра)  $\xi$  – це випадковий елемент з простору  $(M_p(\mathbb{X}), \mathcal{M}_p(\mathbb{X}))$ .*

Зручною є така інтерпретація цього означення: для кожної фіксованої елементарної події  $\xi \in M_p(\mathbb{X})$  локально скінченною точковою мірою на просторі  $\mathbb{X}$ , а для кожної множини  $B \in \mathcal{B}(\mathbb{X})$  її міра  $\xi(B)$  є невід'ємною цілочисельною випадковою величиною – кількістю атомів  $\xi$ , що потрапляють в  $B$ .

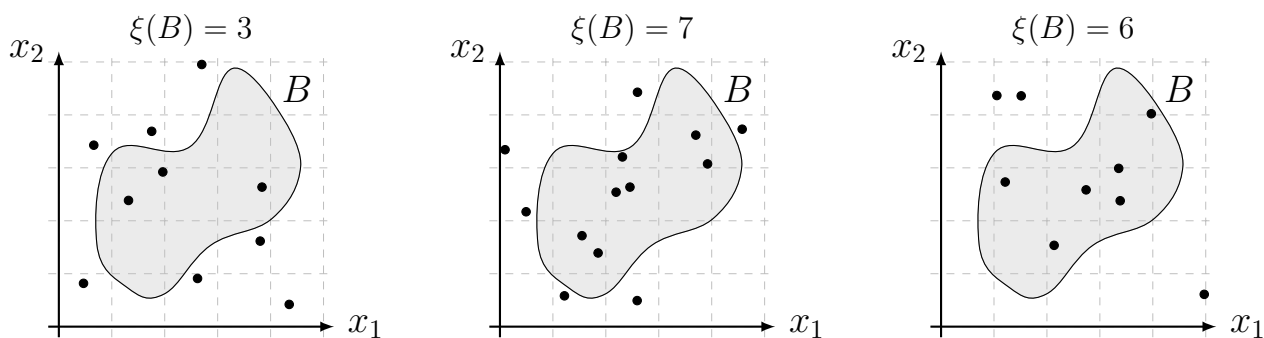


Рисунок 2.1. Приклад реалізацій  $\xi(B)$ .

Для точкових процесів, як і для випадкових мір взагалі, існує кілька типів збіжності (напр., розділ 4 в [10]). Ми зосередимось лише на одному з них.

**Означення 2.** *Послідовність не випадкових мір  $\mu_n$  на  $\mathbb{X}$  грубо збігається до міри  $\mu$ , якщо  $\int_{\mathbb{X}} f d\mu_n \rightarrow \int_{\mathbb{X}} f d\mu$  для всіх обмежених неперервних функцій на  $\mathbb{X}$  з обмеженим носієм. Позначається  $\mu_n \xrightarrow{v} \mu$ .*

**Означення 3.** Послідовність точкових процесів  $\xi_n$  на  $\mathbb{X}$  *грубо збігається за розподілом* до точкового процесу  $\xi$ , якщо виконується одна з трьох еквівалентних умов:

- 1)  $\mathbb{E}\varphi(\xi_n) \rightarrow \mathbb{E}\varphi(\xi)$  для кожного відображення  $\varphi : M_p(\mathbb{X}) \rightarrow \mathbb{R}$ , обмеженого та неперервного в сенсі грубої збіжності;
- 2)  $\int_{\mathbb{X}} f d\xi_n \xrightarrow{d} \int_{\mathbb{X}} f d\xi$  для всіх обмежених неперервних функцій на  $\mathbb{X}$  з обмеженим носієм;
- 3)  $(\xi_n(B_1), \dots, \xi_n(B_m)) \xrightarrow{d} (\xi(B_1), \dots, \xi(B_m))$  для всіх скінченних наборів множин  $B_1, \dots, B_m$  з  $\widehat{\mathcal{B}}_\xi = \{B \in \widehat{\mathcal{B}}(\mathbb{X}) : \mathbb{E}\xi(\partial B) = 0\}$ .

Позначається  $\xi_n \xrightarrow{vd} \xi$ .

Наведемо ознаку грубої збіжності за розподілом (теорема 4.18 в [10]), якою ми будемо користуватись.

**Теорема 1.** Нехай  $\xi_n$  – послідовність точкових процесів на  $\mathbb{X}$ ,  $\xi$  – точковий процес, для якого  $\mathbb{P}(\forall x \in \mathbb{X} : \xi(\{x\}) \leq 1) = 1$ . Нехай також  $\mathcal{U}$  – кільце множин з  $\widehat{\mathcal{B}}_\xi$  таке, що кожна відкрита підмножина  $\mathbb{X}$  є зліченим об'єднанням множин з  $\mathcal{U}$  та кожна множину з  $\widehat{\mathcal{B}}_\xi$  можна покрити скінченною кількістю множин з  $\mathcal{U}$ , а  $\mathcal{I} \subset \mathcal{U}$  – напівкільце з такими ж властивостями. Тоді  $\xi_n \xrightarrow{vd} \xi$ , якщо

- 1)  $\lim_{n \rightarrow \infty} \mathbb{P}(\xi_n(U) = 0) = \mathbb{P}(\xi(U) = 0)$  для всіх  $U \in \mathcal{U}$ ;
- 2)  $\limsup_{n \rightarrow \infty} \mathbb{E}\xi_n(I) \leq \mathbb{E}\xi(I) < \infty$  для всіх  $I \in \mathcal{I}$ .

Наступна теорема (напр., розділ 3.5 в [16]) матиме продуктивне застосування для доведення різних наслідків з основної теореми даної роботи.

**Теорема 2.** Нехай  $\xi_n \xrightarrow{vd} \xi$ , а відображення  $\varphi : M_p(\mathbb{X}) \rightarrow \mathbb{R}^k$  є таким, що  $\mathbb{P}(\varphi \in \text{грубо неперервним в } \xi) = 1$ . Тоді  $\varphi(\xi_n) \xrightarrow{d} \varphi(\xi)$ .

Нарешті, наведемо означення одного з найважливіших прикладів точкових процесів (напр., [11] чи розділ 3.3 в [16]), який є ключовим в роботі.

**Означення 4.** Нехай  $\mu$  – локально скінченна міра на  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ . Процесом Пуассона з мірою інтенсивності  $\mu$  називається точковий процес  $N$ , для якого

- 1) для всіх  $B \in \mathcal{B}(\mathbb{X})$ ,  $N(B) \sim \text{Pois}(\mu(B))$ , якщо  $\mu(B) < \infty$ , і  $N(B) = \infty$  інакше;
- 2) для множин  $B_1, \dots, B_m \in \mathcal{B}(\mathbb{X})$ , що не перетинаються, випадкові величини  $N(B_1), \dots, N(B_m)$  є незалежними.

## 2.2. Основна теорема

Задамо метричний простір, за допомогою якого можна описати структуру перестановки. Для  $k \in \mathbb{N}$  нехай

$$\mathbb{X}_k = \{\mathbf{x} = (x_1, \dots, x_k) \in [0, 1]^k : \min\{x_1, \dots, x_k\} = x_1\}.$$

Остання рівність тут пояснюється тим, що будь-який елемент циклу можна вважати його «початком».

Розглядатимемо простір  $\mathbb{X} = \bigcup_{k=1}^{\infty} \mathbb{X}_k$  з метрикою

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \rho_k(\mathbf{x}_1, \mathbf{x}_2), & \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{X}_k, \\ \sqrt{\max\{k_1, k_2\}}, & \mathbf{x}_1 \in \mathbb{X}_{k_1}, \mathbf{x}_2 \in \mathbb{X}_{k_2}, k_1 \neq k_2, \end{cases}$$

де  $\rho_k$  – евклідова метрика на  $\mathbb{X}_k$ . Нерівність трикутника для  $\rho$  виконується, тому що  $\sup_{\mathbf{x}_1, \mathbf{x}_2} \rho_k(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{k}$ . Зауважимо, що метричний простір  $(\mathbb{X}, \rho)$  є повним та сепарабельним. На ньому також можна описати  $\sigma$ -алгебру  $\mathcal{B}(\mathbb{X})$  за правилом  $(B \in \mathcal{B}(\mathbb{X})) \Leftrightarrow (\forall k : B \cap \mathbb{X}_k \in \mathcal{B}(\mathbb{X}_k))$ .

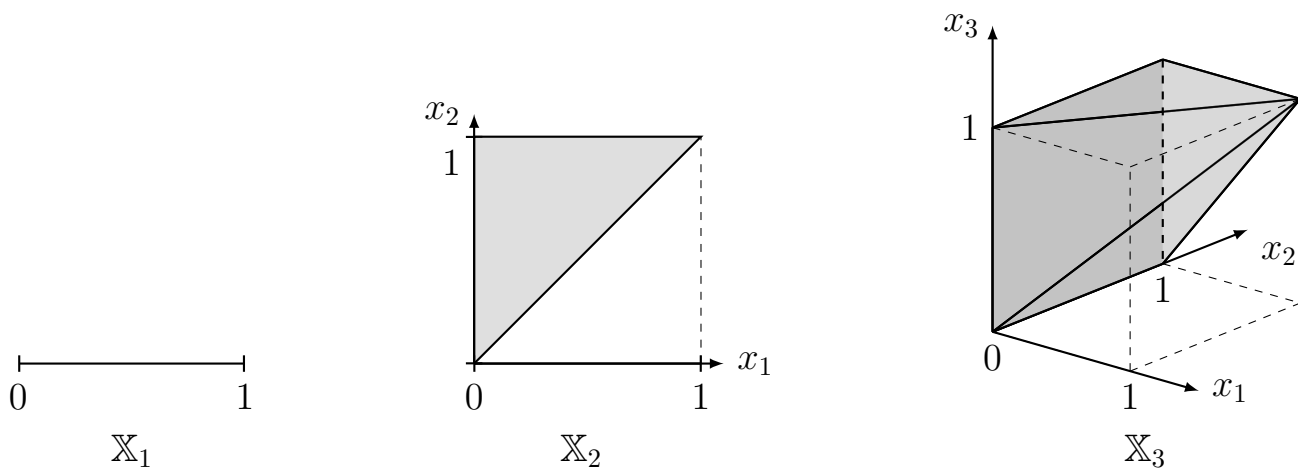


Рисунок 2.2. Перші «поверхи» множини  $\mathbb{X}$ .

Нехай  $\delta_{\mathbf{x}} = \mathbb{1}\{\mathbf{x} \in \cdot\}$  – міра Дірака в  $\mathbf{x}$ . Для перестановки  $\sigma_n$  з розподілом (1) визначимо точковий процес  $\Psi_n$  на  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$  як

$$\Psi_n = \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in [n]}^{\neq} \delta_{\left(\frac{i_1}{n}, \dots, \frac{i_k}{n}\right)} \mathbb{1}\{\sigma_n(i_1) = i_2, \dots, \sigma_n(i_k) = i_1\}, \quad (3)$$

де  $\sum^{\neq}$  означає, що сума береться за попарно різними  $i_1, \dots, i_k \in [n]$ . Оскільки  $\Psi_n$  розглядаються на  $\mathbb{X}$ , то за визначенням внутрішня сума береться за тими наборами

$(i_1, \dots, i_k)$ , де  $i_1$  є найменшим елементом. Таким чином, атоми  $\Psi_n$ , розташовані в точках, що взаємно однозначно відповідають циклам  $\sigma_n$ , повністю описують цю перестановку.

**Теорема 3.** *Нехай послідовність точкових процесів  $\Psi_n$  визначена (3). Якщо виконується  $\lim_{n \rightarrow \infty} \frac{h_{n-1}}{h_n} = 1$ , то  $\Psi_n$  грубо збігається за розподілом до процесу Пуассона  $\Psi$  на  $\mathbb{X}$ , міра інтенсивності якого визначена як  $\lambda(B) = \sum_{k=1}^{\infty} \theta_k \lambda_k(B \cap \mathbb{X}_k)$ , де  $\lambda_k$  – міра Лебега на  $\mathbb{X}_k$ .*

**Зауваження.** Нехай  $\Psi_n^{(k)}$  та  $\Psi^{(k)}$  – звуження на  $\mathbb{X}_k$  процесів  $\Psi_n$  та  $\Psi$ , відповідно. Тоді за означенням грубої збіжності за розподілом маємо  $\Psi_n^{(k)} \xrightarrow{vd} \Psi^{(k)}$  для кожного  $k \in \mathbb{N}$ . Також, за властивістю звуження точкового процесу Пуассона (напр., теорема 5.2 в [11]),  $\Psi^{(k)}$  є незалежними точковими процесами Пуассона з мірами інтенсивності  $\theta_k \lambda_k$ . Таким чином, оскільки  $\Psi_n^{(k)}(\mathbb{X}_k) = C_k(\sigma_n)$ , результати типу (2) безпосередньо впливають з теореми 3.

Перш ніж перейти до доведення теореми, розглянемо одну допоміжну лему. Тут і далі припускатимемо, що виконується  $\lim_{n \rightarrow \infty} \frac{h_{n-1}}{h_n} = 1$ .

**Лема 4.** *Нехай  $r \geq 1$ , а  $\mathbf{i}^{(j)} = (i_1^{(j)}, \dots, i_{k_j}^{(j)})$ ,  $j \in [r]$  – диз'юнктні набори різних натуральних чисел, причому  $\mathbf{i}^{(j)}/n \in \mathbb{X}_{k_j}$ . Тоді*

$$\mathbb{P} \left( \sigma_n \text{ містить цикли } \mathbf{i}^{(1)}, \dots, \mathbf{i}^{(r)} \right) \sim \frac{\prod_{j=1}^r \theta_{k_j}}{n^{k_1 + \dots + k_r}}, \quad n \rightarrow \infty. \quad (4)$$

*Доведення.* Нехай  $I$  – множина всіх елементів  $\mathbf{i}^{(j)}$ ,  $j \in [r]$ ,  $s = \#I = k_1 + \dots + k_r$ , де  $\#$  – кількість елементів. Згідно з (1), ймовірність в (4) дорівнює

$$\begin{aligned} \sum_{\tilde{\pi} \in \mathcal{S}_{[n] \setminus I}} \mathbb{P} \left( \sigma_n = \mathbf{i}^{(1)} \circ \dots \circ \mathbf{i}^{(r)} \circ \tilde{\pi} \right) &= \sum_{\tilde{\pi} \in \mathcal{S}_{[n] \setminus I}} \frac{1}{h_n n!} \prod_{k=1}^{\infty} \theta_k^{\#\{j: k_j=k\} + C_k(\tilde{\pi})} = \\ &= \frac{1}{h_n n!} \prod_{j=1}^r \theta_{k_j} \cdot \sum_{\tilde{\pi} \in \mathcal{S}_{[n] \setminus I}} \prod_{k=1}^{\infty} \theta_k^{C_k(\tilde{\pi})} \stackrel{*}{=} \frac{h_{n-s} (n-s)!}{h_n n!} \prod_{j=1}^r \theta_{k_j} = \frac{h_{n-s}}{h_n} \cdot \frac{\prod_{j=1}^r \theta_{k_j}}{n^{k_1 + \dots + k_r}}, \end{aligned}$$

де рівність  $\stackrel{*}{=}$  пояснюється тим, що другий множник перед цим фактично є сумою ненормованих ймовірностей з розподілу (1), але на множині  $[n] \setminus I$  з  $n-s$  елементів. Залишилось зауважити, що  $\lim_{n \rightarrow \infty} \frac{h_{n-s}}{h_n} = 1$  за припущенням.  $\square$

*Доведення теореми 3.* Скористаємось ознакою збіжності з теореми 1. Достатньо довести, що

- 1)  $\lim_{n \rightarrow \infty} \mathbb{E}\Psi_n(B) = \mathbb{E}\Psi(B)$  для всіх множин виду  $B = \times_{j=1}^k \langle a_j, b_j \rangle \in \mathbb{X}_k$  для довільного  $k \in \mathbb{N}$ , де  $\langle \rangle$  означає ( або  $[, ) -$  чи ];
- 2)  $\lim_{n \rightarrow \infty} \mathbb{P}(\Psi_n(U) = 0) = \mathbb{P}(\Psi(U) = 0)$  для скінченних об'єднань множин такого ж виду (в тому числі з різних  $\mathbb{X}_k$ ).

Нехай  $k \in \mathbb{N}$ ,  $B \in \mathbb{X}_k$  – множина заданого виду, а  $B_{\neq}$  – множина точок  $B$  з попарно різними координатами. Згідно з (3), маємо

$$\mathbb{E}\Psi_n(B) = \sum_{(i_1, \dots, i_k)/n \in B}^{\neq} \mathbb{P}(\sigma_n(i_1) = i_2, \dots, \sigma_n(i_k) = i_1).$$

З леми отримуємо, що всі доданки в правій частині є однаковими та асимптотично еквівалентними  $\theta_k/n^k$ . Отже, при  $n \rightarrow \infty$

$$\mathbb{E}\Psi_n(B) \sim \frac{\theta_k}{n^k} \cdot \#(B_{\neq} \cap (\mathbb{Z}^k/n)) \rightarrow \theta_k \lambda_k(B) = \lambda(B) = \mathbb{E}\Psi(B),$$

щр доводить перше твердження.

Тепер доведемо друге. Нехай  $U$  – множина заданого виду,  $U_m = U \cap \mathbb{X}_m$ . Оскільки  $U_m = \emptyset$  при всіх  $m$ , більших за деяке  $k \geq 1$ , то  $U = \bigcup_{m=1}^k U_m$ . Нехай  $\mathbf{i}_m$  та  $\mathbf{i}_m^{(\cdot)}$  – диз'юнктні набори  $m$  різних натуральних чисел, а  $\bigcirc_{j_m=1}^{r_m} \mathbf{i}_m^{(j_m)}$  позначає композицію  $\mathbf{i}_m^{(1)} \circ \dots \circ \mathbf{i}_m^{(r_m)}$  з  $r_m$  циклів, заданих цими наборами. Для кожного  $R \geq 1$  за нерівністю Бонферроні (напр., [7]) маємо

$$\begin{aligned} \mathbb{P}(\Psi_n(U) = 0) &= 1 - \mathbb{P}\left(\bigcup_{m=1}^k \bigcup_{\mathbf{i}_m/n \in U_m} \{\sigma_n \text{ містить цикл } \mathbf{i}_m\}\right) \leq \\ &\leq \sum_{r_1, \dots, r_k=0}^{2R} (-1)^{r_1 + \dots + r_k} \sum_{\substack{\mathbf{i}_1^{(1)}/n, \dots, \mathbf{i}_1^{(r_1)}/n \in U_1, \\ \vdots \\ \mathbf{i}_k^{(1)}/n, \dots, \mathbf{i}_k^{(r_k)}/n \in U_k}}^* \mathbb{P}\left(\sigma_n \text{ містить } \bigcirc_{j_1=1}^{r_1} \mathbf{i}_1^{(j_1)} \circ \dots \circ \bigcirc_{j_k=1}^{r_k} \mathbf{i}_k^{(j_k)}\right), \quad (5) \end{aligned}$$

де  $\sum^*$  означає, що сума береться за усіма неупорядкованими множинами наборів різних чисел. Також має місце аналогічна оцінка  $\mathbb{P}(\Psi_n(U) = 0)$  знизу, де сума береться за  $r_1$  від 0 до  $2R - 1$ , а за всіма іншими змінними – так само від 0 до  $2R$ .

Згідно з лемою, всі доданки в кожній  $\sum^*$  не залежать від  $\mathbf{i}_m^{(j_m)}$ ,  $j_m = 1, \dots, r_m$ ,  $m = 1, \dots, k$ , і кожен з них дорівнює

$$\frac{h_{n - \sum_{m=1}^k m r_m} (n - \sum_{m=1}^k m r_m)!}{h_n n!} \cdot \prod_{m=1}^k \theta_m^{r_m}.$$



Нехай  $S_n(r_1, \dots, r_k)$  позначає кількість доданків в кожній  $\sum^*$ :

$$S_n(r_1, \dots, r_k) = \sum_{\substack{\mathbf{i}_1^{(1)}/n, \dots, \mathbf{i}_1^{(r_1)}/n \in \mathbb{X}_1, \\ \vdots \\ \mathbf{i}_k^{(1)}/n, \dots, \mathbf{i}_k^{(r_k)}/n \in \mathbb{X}_k}} \frac{\mathbb{1}\{\text{всі набори диз'юнктні}\}}{\prod_{m=1}^k r_m!} \prod_{m=1}^k \mathbb{1}\{\mathbf{i}_m^{(1)}/n, \dots, \mathbf{i}_m^{(r_m)}/n \in U_m\},$$

де ділення на  $\prod_{m=1}^k r_m!$  пояснюється тим, що в означенні  $S_n$  сума береться за впорядкованими множинами наборів. Отже, права частина нерівності (5) дорівнює

$$\sum_{r_1, \dots, r_k=0}^{2R} (-1)^{r_1+\dots+r_k} \cdot \frac{h_{n-\sum_{m=1}^k mr_m} (n - \sum_{m=1}^k mr_m)!}{h_n n!} \cdot \prod_{m=1}^k \theta_m^{r_m} \cdot S_n(r_1, \dots, r_k),$$

Зауважимо, що  $\frac{S_n(r_1, \dots, r_k)}{n^{r_1+\dots+kr_k}}$  можна розглядати як інтегральну суму для

$$\int_{U_1^{r_1} \times \dots \times U_k^{r_k}} \frac{\mathbb{1}\{\text{всі елементи всіх } \mathbf{x}_m^{(j_m)} \text{ є різними}\}}{\prod_{m=1}^k r_m!} \prod_{m=1}^k \prod_{j_m=1}^{r_m} d\mathbf{x}_m^{(j_m)} = \prod_{m=1}^k \frac{(\lambda_m(U_m))^{r_m}}{r_m!}.$$

Таким чином, оскільки  $\frac{(n-\sum_{m=1}^k mr_m)!}{n!} \sim n^{\sum_{m=1}^k mr_m}$  та  $\lim_{n \rightarrow \infty} \frac{h_{n-1}}{h_n} = 1$ , то кожен доданок в отриманій оцінці збігається при  $n \rightarrow \infty$  до  $(-1)^{r_1+\dots+r_k} \prod_{m=1}^k \frac{(\theta_m \lambda_m(U_m))^{r_m}}{r_m!}$ .

Таким чином, згадуючи про аналогічну оцінку знизу, отримуємо

$$\begin{aligned} \sum_{r_1=1}^{2R-1} \sum_{r_2, \dots, r_k=0}^{2R} (-1)^{r_1+\dots+r_k} \prod_{m=1}^k \frac{(\theta_m \lambda_m(U_m))^{r_m}}{r_m!} &\leq \lim_{n \rightarrow \infty} \mathbb{P}(\Psi_n(U) = 0) \leq \\ &\leq \sum_{r_1, \dots, r_k=0}^{2R} (-1)^{r_1+\dots+r_k} \prod_{m=1}^k \frac{(\theta_m \lambda_m(U_m))^{r_m}}{r_m!}. \end{aligned}$$

Оскільки  $R$  є довільним, можемо спрямувати  $R \rightarrow \infty$ , щоб отримати

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\Psi_n(U) = 0) &= \sum_{r_1, \dots, r_k=0}^{\infty} (-1)^{r_1+\dots+r_k} \prod_{m=1}^k \frac{(\theta_m \lambda_m(U_m))^{r_m}}{r_m!} = \\ &= \exp \left\{ - \sum_{m=1}^k \theta_m \lambda_m(U_m) \right\} = \exp \{-\lambda(U)\} = \mathbb{P}(\Psi(U) = 0). \end{aligned}$$

Таким чином, теорему доведено повністю. □

### 2.3. Наслідки та приклади застосування

Перш за все, звернемо увагу на те, що означення збіжності  $\Psi_n \xrightarrow{vd} \Psi$  де-що спрощуються, якщо врахувати побудову  $\mathbb{X}$  та  $\Psi$ . Так, з огляду на топологію простору  $\mathbb{X}$ , кожна збіжна послідовність  $\{\mathbf{x}_n\} \subset \mathbb{X}$ , починаючи з певного номера, належить лише одному «поверху»  $\mathbb{X}_k$ , а обмежені множини мають непорожній перетин лише зі скінченною кількістю різних  $\mathbb{X}_k$ . Це означає, що функції виду  $f(\mathbf{x}) = \sum_{k \in \mathcal{K}} f_k(\mathbf{x}) \mathbb{1}\{\mathbf{x} \in \mathbb{X}_k\}$ , де  $\mathcal{K}$  – деяка скінченна підмножина  $\mathbb{N}$ , а  $f_k(\mathbf{x}) : \mathbb{X}_k \rightarrow \mathbb{R}$  – неперервні, є неперервними обмеженими функціями на  $\mathbb{X}$  з обмеженим носієм. Також, враховуючи побудову міри  $\lambda$ , умова  $\mathbb{E}\Psi(\partial B) = 0$  означає, що  $\lambda_k(\partial B \cap \mathbb{X}_k) = 0$  для всіх  $k$ .

Таким чином, теорема 3 дозволяє отримувати граничні теореми для випадкових величин, пов'язаних зі складом циклів випадкової перестановки. Особливо цікавими є випадки, коли можна описати розподіл граничної випадкової величини, користуючись певними властивостями процесу Пуассона. Як і раніше, припускатимемо, що виконується  $\lim_{n \rightarrow \infty} \frac{h_{n-1}}{h_n} = 1$ .

**Твердження 1.** Нехай  $k \in \mathbb{N}$ ,  $f_k : \mathbb{X}_k \rightarrow \mathbb{R}$ ,  $k \in [m]$  – неперервні функції. Тоді

$$\left( \int_{\mathbb{X}_1} f_1 d\Psi_n^{(1)}, \dots, \int_{\mathbb{X}_m} f_m d\Psi_n^{(m)} \right) \xrightarrow{d} \left( \int_{\mathbb{X}_1} f_1 d\Psi^{(1)}, \dots, \int_{\mathbb{X}_m} f_m d\Psi^{(m)} \right),$$

причому випадкові величини  $J_k = \int_{\mathbb{X}_k} f_k d\Psi^{(k)}$  є незалежними. Перетворення Лапласа величин  $J_k$  можна записати як

$$\mathbb{E}e^{-tJ_k} = \exp \left\{ -\theta_k \int_{\mathbb{X}_k} \left( 1 - e^{-tf_k(\mathbf{x})} \right) d\mathbf{x} \right\}, \quad t \geq 0.$$

*Доведення.* Згідно з теоремою Крамера-Вольда (напр., розділ 29 в [4]), для доведення збіжності достатньо довести, що

$$\forall (t_1, \dots, t_m) \in \mathbb{R}^m : \sum_{k=1}^m t_k \int_{\mathbb{X}_k} f_k d\Psi_n^{(k)} \xrightarrow{d} \sum_{k=1}^m t_k \int_{\mathbb{X}_k} f_k d\Psi^{(k)}.$$

Цей факт є наслідком лінійності інтегралів та збіжності  $\int_{\mathbb{X}} f d\Psi_n \xrightarrow{d} \int_{\mathbb{X}} f d\Psi$  для  $f(\mathbf{x}) = \sum_{k=1}^m t_k f_k(\mathbf{x}) \mathbb{1}\{\mathbf{x} \in \mathbb{X}_k\}$ . Незалежність  $J_1, \dots, J_m$  є наслідком незалежності  $\Psi^{(1)}, \dots, \Psi^{(m)}$ . Існування перетворення Лапласа для  $J_k$  впливає з обмеженості  $f_k$  як неперервної функції на компактi, а його вигляд — з властивостей функціоналу Лапласа процесу Пуассона (напр., твердження 3.6 в [16]).  $\square$

Наведемо приклад застосування твердження 1. Нехай  $S_n^{(k)}$  – сума елементів всіх циклів довжини  $k$  в  $\sigma_n$ .

**Твердження 2.** Для визначених вище  $S_n^{(k)}$ :

- 1)  $\frac{S_n^{(k)}}{n} \xrightarrow{d} S^{(k)}$  при  $n \rightarrow \infty$ , де  $\mathbb{E}e^{-tS^{(k)}} = \exp \left\{ \frac{\theta_k}{k} \left( \left( \frac{1-e^{-t}}{t} \right)^k - 1 \right) \right\}$ ,  $t > 0$ ;
- 2)  $\mathbb{P}(S^{(1)} \leq x) = e^{-\theta_1} \sum_{j=0}^{\lfloor x \rfloor} (-1)^j \frac{1}{j!} (\theta_1(x-j))^{\frac{j}{2}} I_j \left( 2\sqrt{\theta_1(x-j)} \right)$  для  $x \geq 0$ , де  $I_j$  – модифікована функція Бесселя 1 роду (напр., §10.25(ii) в [13]).

*Доведення.* Для доведення збіжності в 1) застосуємо твердження 1 для функцій  $f_k : \mathbb{X}_k \rightarrow [0, +\infty)$ , визначених як  $f_k(\mathbf{x}) = x_1 + \dots + x_k$ . Оскільки  $f_k$  є симетричними, то

$$\int_{\mathbb{X}_k} (1 - e^{-tf_k(\mathbf{x})}) d\mathbf{x} = \frac{1}{k} \int_{[0,1]^k} (1 - e^{-tf_k(\mathbf{x})}) d\mathbf{x} = \frac{1}{k} \left( 1 - \left( \int_0^1 e^{-tx} dx \right)^k \right),$$

звідки й отримуємо вид перетворення Лапласа для  $S^{(k)}$ .

Перейдемо до доведення 2). Оскільки  $\int_s^\infty te^{-tx} dx = e^{-ts}$ , то

$$\mathbb{E}e^{-tS^{(1)}} = \mathbb{E} \int_0^\infty te^{-tx} \mathbf{1}\{S^{(1)} \leq x\} dx = t \int_0^\infty e^{-tx} \mathbb{P}(S^{(1)} \leq x) dx.$$

Таким чином, перетворення Лапласа функції розподілу  $\mathbb{P}(S^{(1)} \leq x)$  дорівнює

$$G(t) = t^{-1} \mathbb{E}e^{-tS^{(1)}} = t^{-1} e^{-\theta_1} e^{\theta_1/t} e^{-\theta_1 e^{-t}/t} = \sum_{j=0}^{\infty} \frac{(-\theta_1)^j}{j!} e^{-\theta_1} \underbrace{t^{-j-1} e^{\theta_1/t} e^{-jt}}_{G_j(t)}, \quad t > 0.$$

З таблиць перетворення Лапласа спеціальних функцій (напр., 5.5.35 в [18]) та властивості запізнення оригіналу отримуємо обернене перетворення  $G_j(t)$ :

$$F_j(x) = (\theta_1(x-j))^{j/2} I_j \left( 2\sqrt{\theta_1(x-j)} \right) \mathbf{1}\{x \geq j\}.$$

Це означає, що для всіх  $t > 0$  та  $R \in \mathbb{N}$

$$\int_0^\infty e^{-tx} \sum_{j=0}^R \frac{(-\theta_1)^j}{j!} e^{-\theta_1} F_j(x) dx = \sum_{j=0}^R \frac{(-\theta_1)^j}{j!} e^{-\theta_1} G_j(t).$$

Оскільки  $I_j(x)$  спадає за  $j$  та зростає за  $x$  (напр., §10.37 в [13]), а  $I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}}$

при  $x \rightarrow \infty$  (там само, §10.30(ii)), то

$$\left| \sum_{j=0}^R \frac{(-\theta_1)^j}{j!} F_j(x) \right| \leq I_0(2\sqrt{\theta_1 x}) \sum_{j=0}^{\infty} \frac{1}{j!} (\theta_1 x)^{j/2} \sim \frac{e^{3\sqrt{\theta_1 x}}}{2\pi^{1/2} (\theta_1 x)^{1/4}}, \quad x \rightarrow \infty.$$

Таким чином, можна застосувати теорему Лебега про мажоровану збіжність при  $R \rightarrow \infty$ , щоб отримати обернене перетворення Лапласа для  $G(t)$ , яке і буде шуканою функцією розподілу.  $\square$

Перейдемо до випадкових величин іншого виду. За допомогою теореми 2 можна отримувати твердження типу  $\varphi(\Psi_n) \xrightarrow{d} \varphi(\Psi)$ , де  $\varphi : M_p(\mathbb{X}) \rightarrow \mathbb{R}$  – деяке грубо неперервне відображення. Корисною для перевірки грубої неперервності є наступна теорема (твердження 3.13 в [16]).

**Теорема 5.** Нехай  $\mu_n$  – послідовність точкових мір на  $\mathbb{X}$ ,  $\mu_n \xrightarrow{v} \mu$ . Тоді для кожного компакта  $K \subset \mathbb{X}$ , для якого  $\mu(\partial K) = 0$ , існує таке  $N \in \mathbb{N}$  та така нумерація атомів  $\mu_n$  та  $\mu$ , що при всіх  $n \geq N$

$$\mu_n(\cdot \cap K) = \sum_{i=1}^m \delta_{x_i^{(n)}}, \quad \mu(\cdot \cap K) = \sum_{i=1}^m \delta_{x_i},$$

і при цьому  $x_i^{(n)} \rightarrow x_i$  для всіх  $i$ .

Наведемо кілька прикладів таких тверджень. Надалі позначатимемо множини атомів  $\Psi_n^{(k)}$  та  $\Psi^{(k)}$  як  $\mathcal{A}_n^{(k)}$  та  $\mathcal{A}^{(k)}$  відповідно.

Для  $k \geq 2$  назвемо *розмахом циклу* довжини  $k$  різницю між його найбільшим та найменшим елементами. Нехай  $r_n^{(k)}$  та  $R_n^{(k)}$  – найменший та найбільший розмахи серед усіх циклів довжини  $k$  в  $\sigma_n$  відповідно, причому  $r_n^{(k)} = n$ ,  $R_n^{(k)} = 0$ , якщо таких циклів немає. Наприклад, для  $\sigma = (1, 5, 6) \circ (2, 9, 4) \circ (3, 8, 7) \in \mathcal{S}_9$  маємо  $r_9^{(3)} = \min\{6 - 1, 9 - 2, 8 - 3\} = 5$ ,  $R_9^{(3)} = \max\{6 - 1, 9 - 2, 8 - 3\} = 7$ ,  $r_9^{(k)} = 0$  та  $R_9^{(k)} = 0$  при  $k \neq 3$ .

**Твердження 3.** Для  $r_n^{(k)}$  та  $R_n^{(k)}$  виконується  $\frac{r_n^{(k)}}{n} \xrightarrow{d} r^{(k)}$  та  $\frac{R_n^{(k)}}{n} \xrightarrow{d} R^{(k)}$  при  $n \rightarrow \infty$ , де

$$\mathbb{P}(r^{(k)} \leq x) = 1 - \exp \left\{ -\frac{\theta_k}{k} (kx^{k-1} - (k-1)x^k) \right\}, \quad x \in [0, 1),$$

$$\mathbb{P}(R^{(k)} \leq x) = \exp \left\{ \frac{\theta_k}{k} (kx^{k-1} - (k-1)x^k - 1) \right\}, \quad x \in [0, 1).$$

*Доведення.* Нехай  $f_k : \mathbb{X}_k \rightarrow [0, 1]$  задано як  $f_k(\mathbf{x}) = \max\{x_2, \dots, x_k\} - x_1$ . Тоді  $\frac{r_n^{(k)}}{n} = \min_{\mathbf{x} \in \mathcal{A}_n^{(k)}} f_k(\mathbf{x})$ , причому відповідне відображення  $M_p(\mathbb{X}_k) \rightarrow [0, 1]$  є грубо неперервним за теоремою 5. Отже,  $\frac{r_n^{(k)}}{n} \xrightarrow{d} \min_{\mathbf{x} \in \mathcal{A}^{(k)}} f_k(\mathbf{x}) = r^{(k)}$ . Для того, щоб отримати функцію розподілу, запишемо її як

$$\begin{aligned} \mathbb{P}\left(r^{(k)} \leq x\right) &= 1 - \mathbb{P}\left(\Psi^{(k)}\{\mathbf{x} \in \mathbb{X}_k : f_k(\mathbf{x}) \leq x\} = 0\right) = \\ &= 1 - \exp\left\{-\theta_k \lambda_k \{\mathbf{x} \in \mathbb{X}_k : f_k(\mathbf{x}) \leq x\}\right\}. \end{aligned}$$

Міру Лебега множини в цьому виразі нескладно обчислити:

$$\int_0^1 \left( \int_{x_1}^{\min\{x_1+x, 1\}} dx_2 \dots dx_k \right) dx_1 = x^{k-1}(1-x) + \frac{x^k}{k}.$$

Доведення для  $R^{(k)} = \max_{\mathbf{x} \in \mathcal{A}^{(k)}} f_k(\mathbf{x})$  аналогічне.  $\square$

Наостанок розглянемо набір випадкових величин, пов'язаних з циклами довжини 1 в  $\sigma_n$ , елементи яких ще називають нерухомими точками.

Нехай  $m_n$  – найменша нерухома точка ( $n+1$ , якщо їх немає),  $M_n$  – найбільша нерухома точка ( $0$ , якщо їх немає),  $\delta_n$  та  $\Delta_n$  – найменший та найбільший спейсинги (відстані між сусідніми нерухомими точками як елементами  $[0, n]$ ), враховуючи два крайніх розміром  $m_n$  та  $n+1 - M_n$ .

Наприклад, для перестановки  $(2) \circ (3) \circ (6) \circ (1, 4, 5) \in \mathcal{S}_6$  маємо  $m_6 = 2$ ,  $M_6 = 6$ ,  $\delta_6 = \min\{2, 3-2, 6-3, 7-6\} = 1$ ,  $\Delta_6 = \max\{2, 3-2, 6-3, 7-6\} = 3$ .

**Твердження 4.**  $\left(\frac{m_n}{n}, \frac{M_n}{n}, \frac{\delta_n}{n}, \frac{\Delta_n}{n}\right) \xrightarrow{d} (m, M, \delta, \Delta)$ , де

$$\mathbb{P}(m \leq x) = 1 - e^{-\theta_1 x}, \quad \mathbb{P}(M \leq x) = e^{\theta_1(x-1)}, \quad x \in [0, 1),$$

$$\delta \stackrel{d}{=} \frac{X_{\nu+1}}{(\nu+1) \sum_{i=1}^{\nu+1} X_i}, \quad \Delta \stackrel{d}{=} \frac{\sum_{i=1}^{\nu+1} X_i/i}{\sum_{i=1}^{\nu+1} X_i}$$

для незалежних  $\nu \sim \text{Pois}(\theta_1)$ ,  $X_i \sim \text{Exp}(1)$ .

*Доведення.* В першу чергу, побудуємо відповідне грубо неперервне відображення  $\varphi : M_p(\mathbb{X}_1) \rightarrow \mathbb{R}^4$ . Нехай

$$\overline{\mathcal{A}}_n^{(1)} = \mathcal{A}_n^{(1)} \cup \{0, 1\}, \quad \mathcal{D}_n = \left\{ (x, y) \in \overline{\mathcal{A}}_n^{(1)} \times \overline{\mathcal{A}}_n^{(1)} : x < y, \Psi_n^{(1)}((x, y)) = 0 \right\},$$

де умова в  $\mathcal{D}_n$  обирає сусідні атоми, що відповідають нерухомим точкам, а  $\mathcal{D}$  –

аналогічна множина для  $\Psi^{(1)}$ . Тоді можемо записати  $\varphi\left(\Psi_n^{(1)}\right)$  як

$$\left(\frac{m_n}{n}, \frac{M_n}{n}, \frac{\delta_n}{n}, \frac{\Delta_n}{n}\right) = \left(\min_{x \in \mathcal{A}_n^{(1)}} x, \max_{x \in \mathcal{A}_n^{(1)}} x, \min_{(x,y) \in \mathcal{D}_n} |x - y|, \max_{(x,y) \in \mathcal{D}_n} |x - y|\right)$$

і скористатися теоремами 5 та 2 для доведення збіжності.

Залишилось отримати явні граничні розподіли. Для  $m$  та  $M$  маємо

$$\begin{aligned} \mathbb{P}(m \leq x) &= 1 - \mathbb{P}\left(\Psi^{(1)}([0, x]) = 0\right) = 1 - e^{-\theta_1 x}, \\ \mathbb{P}(M \leq x) &= \mathbb{P}\left(\Psi^{(1)}((x, 1]) = 0\right) = e^{-\theta_1(1-x)}. \end{aligned}$$

За властивостями  $\Psi^{(1)}$  як однорідного процесу Пуассона, розподіл  $\delta$  (відповідно,  $\Delta$ ) за умови  $\Psi^{(1)}(\mathbb{X}_1) = r$  дорівнює розподілу найменшого (найбільшого) спейсингу  $d_r$  ( $D_r$ ) для  $r$  незалежних випадкових величин, рівномірно розподілених на  $[0, 1]$  (напр., твердження 3.8 в [11]). З теорії спейсингів рівномірного розподілу відомо (напр., [9]), що

$$d_r \stackrel{d}{=} \frac{\min_i Y_i}{\sum_{i=1}^{r+1} Y_i}, \quad D_r \stackrel{d}{=} \frac{\max_i Y_i}{\sum_{i=1}^{r+1} Y_i},$$

де  $Y_1, \dots, Y_{r+1}$  незалежні та мають розподіл  $\text{Exp}(1)$ . Нехай  $Y_{(i)}$ ,  $i \in [r+1]$  – їхні порядкові статистики,  $Y_{(0)} = 0$ , а  $\tau_i = Y_{(i)} - Y_{(i-1)}$ . За теоремою Сухатме-Ренї (напр., теорема 4.6.1 в [5]),  $\tau_i$  є незалежними і  $\tau_i \sim \text{Exp}(r - i + 2)$ , причому

$$\min_i Y_i = \tau_i, \quad \max_i Y_i = \sum_{i=1}^{r+1} \tau_i, \quad \sum_{i=1}^{r+1} Y_i = \sum_{i=1}^{r+1} (r - i + 2)\tau_i.$$

Позначивши  $X_{r-i+2} = (r - i + 2)\tau_i \sim \text{Exp}(1)$ , можемо тепер записати

$$d_r \stackrel{d}{=} \frac{X_{r+1}}{(r+1) \sum_{i=1}^{r+1} X_i}, \quad D_r \stackrel{d}{=} \frac{\sum_{i=1}^{r+1} X_i/i}{\sum_{i=1}^{r+1} X_i}.$$

Оскільки ці рівності виконуються для кожного  $r \geq 0$ , то згадані рівності для  $\delta$  та  $\Delta$  отримуємо, повернувшись від умовних розподілів до безумовного.  $\square$

## РОЗДІЛ 3

### ЧИСЕЛЬНЕ МОДЕЛЮВАННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Насамкінець, наведемо деякі демонстрації отриманих результатів, створені за допомогою чисельного моделювання. Варто зауважити, що на момент написання даної роботи нам не вдалось знайти алгоритмів, за якими можна генерувати випадкові перестановки з розподілом (1) для довільних параметрів  $\theta_k$ , тому в цьому розділі ми обмежимося лише випадком  $\theta_k \equiv \theta > 0$  – перестановками Юенса. Для дослідження величин, пов’язаних з циклами однієї довжини, цього буде достатньо. В цьому випадку формула (1) дещо спрощується (напр., [17]):

$$\mathbb{P}(\sigma_n = \sigma) = \frac{1}{h_n n!} \prod_{k=1}^{\infty} \theta^{C_k(\sigma)} = \frac{1}{h_n n!} \theta^{\sum_{k=1}^{\infty} C_k(\sigma)}, \quad \sigma \in \mathcal{S}_n, \quad (6)$$

де  $h_n = \theta(\theta + 1)\dots(\theta + n - 1)/n!$ , а  $\sum_{k=1}^{\infty} C_k(\sigma)$  є кількістю циклів в  $\sigma$ . Надалі будемо використовувати позначення  $\sigma_n \sim \text{ESF}(n, \theta)$  для перестановок, що мають розподіл (6).

Отже, почнемо з короткого опису алгоритму, який можна використовувати для генерування перестановок Юенса. З його детальним описом та обґрунтуванням можна ознайомитись, наприклад, в [14].

#### 3.1. Процес китайського ресторану

Одразу зауважимо, що даний метод генерування дозволяє отримати не просто перестановку, а й одразу її розклад в композицію незалежних циклів.

Розглянемо випадкові величини  $A_1, A_2, \dots$  з розподілами

$$\mathbb{P}(A_i = j) = \begin{cases} \frac{\theta}{\theta+i-1}, & j = i, \\ \frac{1}{\theta+i-1}, & j \in [i-1]. \end{cases}$$

Перший незалежний цикл починається з 1. Число 2 додається до цього циклу справа, перетворюючи його на  $(1, 2)$  з ймовірністю  $\frac{1}{\theta+1}$ , або ж починає новий цикл з ймовірністю  $\frac{\theta}{\theta+1}$ . Нехай перші  $k-1$  натуральних чисел вже розставлені в цикли. Тоді  $k$  або починає новий цикл з ймовірністю  $\mathbb{P}(A_k = k) = \frac{\theta}{\theta+k-1}$ , або додається справа від  $j$  у вже наявний цикл з ймовірністю  $\mathbb{P}(A_k = j) = \frac{1}{\theta+k-1}$ ,  $j \in [k-1]$ . З цього алгоритму випливає, що ймовірність отримати перестановку на  $[n]$  з  $k$  циклами дорівнює  $\frac{\theta^{k-1}}{(\theta+1)\dots(\theta+n-1)} = \frac{\theta^k}{\theta(\theta+1)\dots(\theta+n-1)}$ , як і в формулі (6).

Цикли, отримані за цим алгоритмом, впорядковані наступним чином: перший

містить 1, другий – найменше число, яке не ввійшло в перший, і так далі.

Варто також зауважити, що цей алгоритм дозволяє отримати не просто випадкову перестановку з розподілом  $\text{ESF}(n, \theta)$ , а послідовність перестановок з розподілами  $\text{ESF}(1, \theta), \dots, \text{ESF}(n, \theta)$ , причому два числа, що в якийсь момент опинилися в одному циклі, й надалі залишаються в ньому ж.

Розглянемо приклад для  $n = 3$ , який можна проілюструвати так:

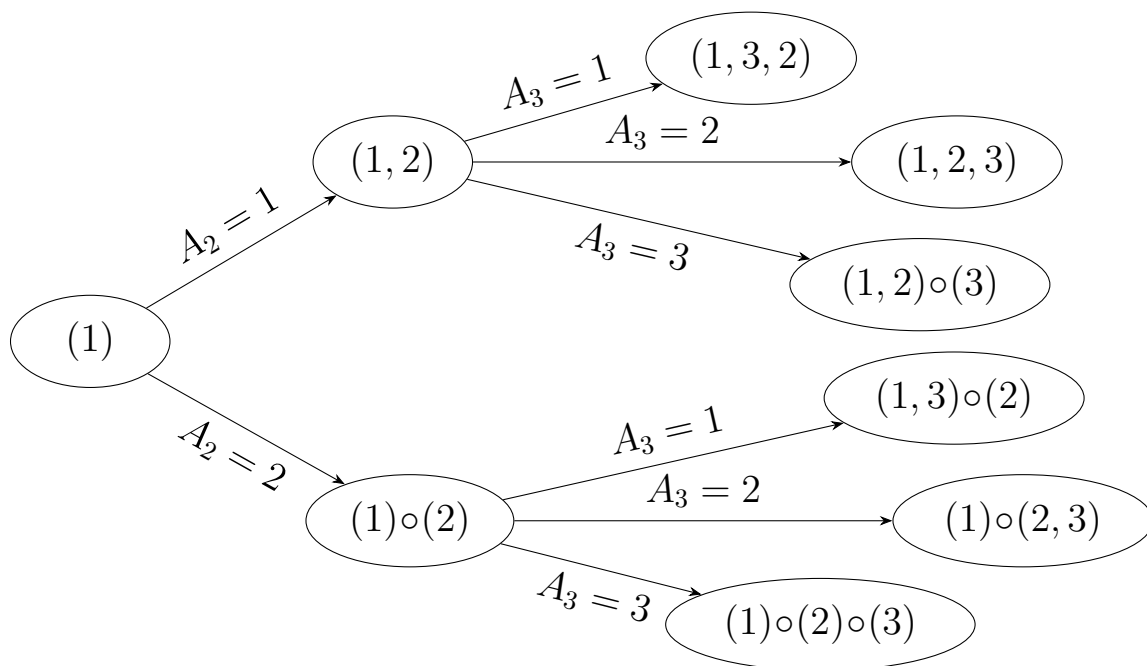


Рисунок 3.1. Процес китайського ресторану для  $n = 3$ .

З цієї діаграми видно, що ймовірність отримати перестановку  $(1) \circ (2, 3)$  обчислюється як  $\mathbb{P}(A_3 = 2) \cdot \mathbb{P}(A_2 = 2) = \frac{1}{\theta+2} \cdot \frac{\theta}{\theta+1}$ . Аналогічно можна перевірити, що ймовірність отримати інші перестановки з двома циклами теж дорівнює  $\frac{1}{\theta+2} \cdot \frac{\theta}{\theta+1}$ , перестановку  $(1, 2, 3)$  з одним циклом –  $\frac{1}{\theta+2} \cdot \frac{1}{\theta+1}$ , а тотожну перестановку  $(1) \circ (2) \circ (3)$  з трьома циклами –  $\frac{\theta}{\theta+2} \cdot \frac{\theta}{\theta+1}$ .

### 3.2. Моделювання результатів

Скористаємося процесом китайського ресторану для моделювання розподілів випадкових величин, пов'язаних з  $\Psi_n$  для  $\sigma_n \sim \text{ESF}(n, \theta)$ , та порівняємо їх з відомими розподілами граничних випадкових величин, які можна отримати з доведених раніше теорем та тверджень.

Для дискретних випадкових величин будемо розглядати емпіричні полігони розподілу, а для інших — емпіричні функції розподілу.



### 3.2.1. Розподіл кількості циклів

Нехай  $U_j \in \mathbb{X}_j$ ,  $j \in [k]$ . Наслідком теореми 3 буде збіжність

$$\Psi_n \left( \bigcup_{j=1}^k U_j \right) = \sum_{j=1}^k \Psi_n^{(j)}(U_j) \xrightarrow{d} \sum_{j=1}^k \Psi^{(j)}(U_j) \sim \text{Pois} \left( \theta \sum_{j=1}^k \lambda_j(U_j) \right).$$

Для спрощення позначимо  $X_n$  початкову випадкову величину зліва, а  $X$  – граничну справа. Перевіримо це для довільного набору множин  $U_j$ . Нехай

$$\begin{aligned} U_1 &= (0.65, 0.85), \\ U_2 &= ((0.25, 0.4) \times (0.5, 0.7)) \cup ((0.05, 0.1) \times (0.15, 0.2)), \\ U_3 &= (0.2, 0.3) \times (0.9, 1) \times (0.35, 0.45). \end{aligned}$$

В цьому випадку  $\lambda_1(U_1) = 0.2$ ,  $\lambda_2(U_2) = 0.0325$ ,  $\lambda_3(U_3) = 0.001$ , тому гранична випадкова величина матиме розподіл  $\text{Pois}(0.2335 \cdot \theta)$ . Нехай  $\theta = 10$ . Порівняємо полігони розподілу для різних  $n$ , використовуючи вибірки розміром  $N = 5000$ .

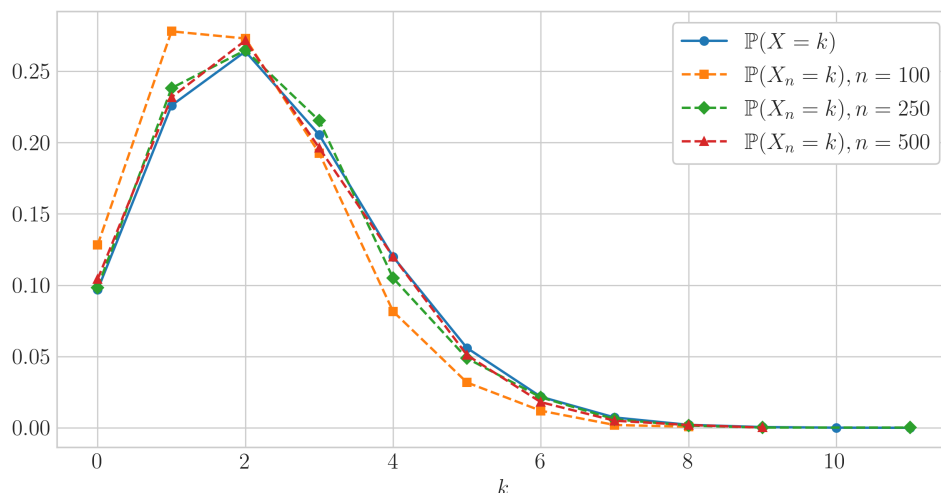


Рисунок 3.2. Полігони розподілу  $X$  та  $X_n$  для  $n = 100, 250, 500$ .

### 3.2.2. Розмахи циклів

Продемонструємо результат твердження 3. Для цього для різних  $n$  згенеруємо  $N = 1000$  перестановок  $\sigma_n$  з розподілом  $\text{ESF}(n, \theta)$ , обчислимо емпіричні функції розподілу  $r_n^{(k)}$  та  $R_n^{(k)}$  та порівняємо їх з відомими функціями розподілу  $r^{(k)}$  та  $R^{(k)}$ . Варто зауважити, що циклів довжини  $k$  в  $\sigma_n$  може не бути взагалі, причому

$$\lim_{n \rightarrow \infty} \mathbb{P}(C_k(\sigma_n) = 0) = \mathbb{P}(\Psi^{(k)}(\mathbb{X}_k) = 0) = e^{-\theta/k}.$$

Через це у функції розподілу  $r^{(k)}$  та  $R^{(k)}$  будуть стрибки величиною  $e^{-\theta/k}$  в точках 1 та 0, відповідно. Для прикладу візьмемо  $\theta = 3$  та  $k = 4$ .

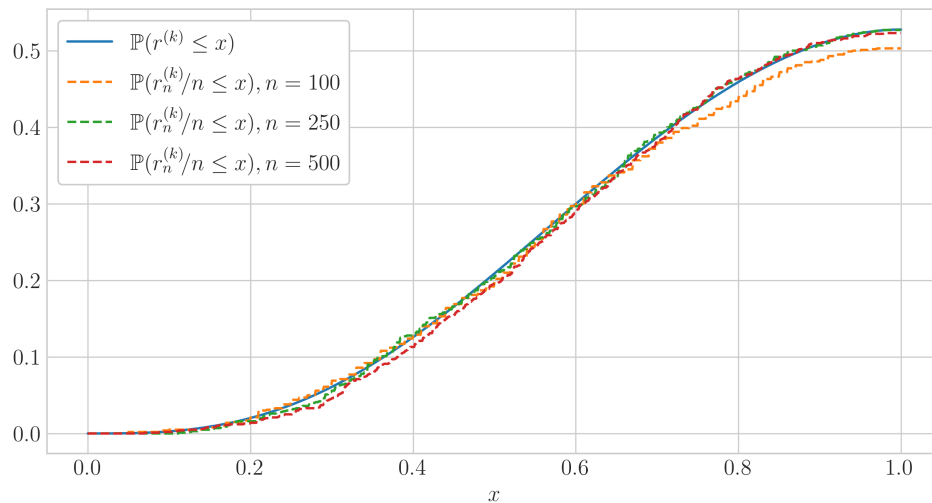


Рисунок 3.3. Функції розподілу  $r^{(k)}$  та  $r_n^{(k)}/n$  для  $n = 100, 250, 500$ .

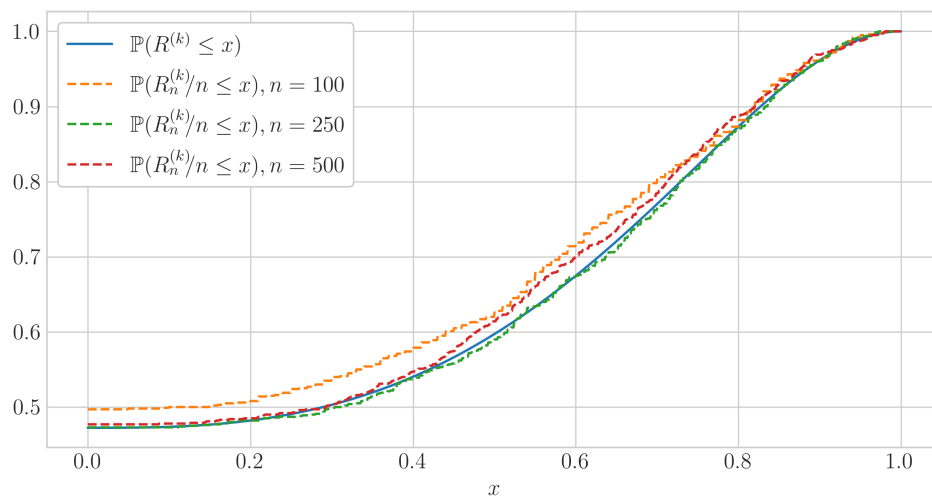


Рисунок 3.4. Функції розподілу  $R^{(k)}$  та  $R_n^{(k)}/n$  для  $n = 100, 250, 500$ .

### 3.2.3. Статистики нерухомих точок

Насамкінець, змодельюємо випадкові величини, пов'язані з нерухомими точками –  $S_n^{(1)}$ ,  $m_n$ ,  $M_n$ ,  $\delta_n$ ,  $\Delta_n$ , які було розглянуто в твердженнях 2 та 4. Нагадаємо, що для відповідних граничних випадкових величин  $S^{(1)}$ ,  $m$ ,  $M$  було отримано функції розподілу в явному вигляді, а для  $\delta$  та  $\Delta$  – лише деякі інші випадкові величини з такими ж розподілами. З огляду на це, для порівняння використовуватимемо емпіричні функції розподілу  $\delta$  та  $\Delta$ , для генерування яких необхідні експоненційні та пуассонівські випадкові величини.

Для суми нерухомих точок  $S_n^{(1)}$  оберемо  $\theta = 5$ . Зауважимо, що у функції розподілу  $S^{(1)}$  в нулі є стрибок величиною  $e^{-\theta} \approx 0.0067$ . Для найменшої та найбільшої

нерухомих точок візьмемо  $\theta = 2$ . У функції розподілу  $m$  є стрибок в точці 1, а у функції розподілу  $M$  – в точці 0, обидва величиною  $e^{-\theta} \approx 0.135$ .

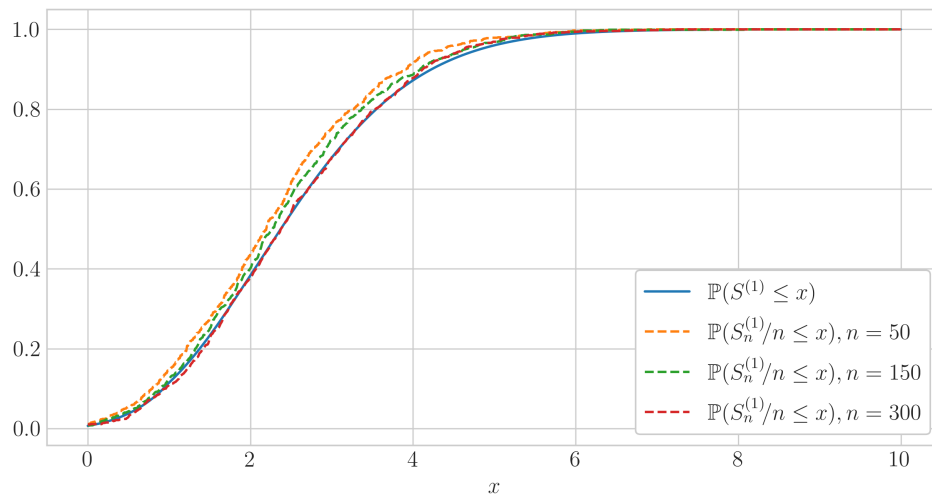


Рисунок 3.5. Функції розподілу  $S^{(1)}$  та  $S_n^{(1)}/n$  для  $n = 50, 150, 300$ .

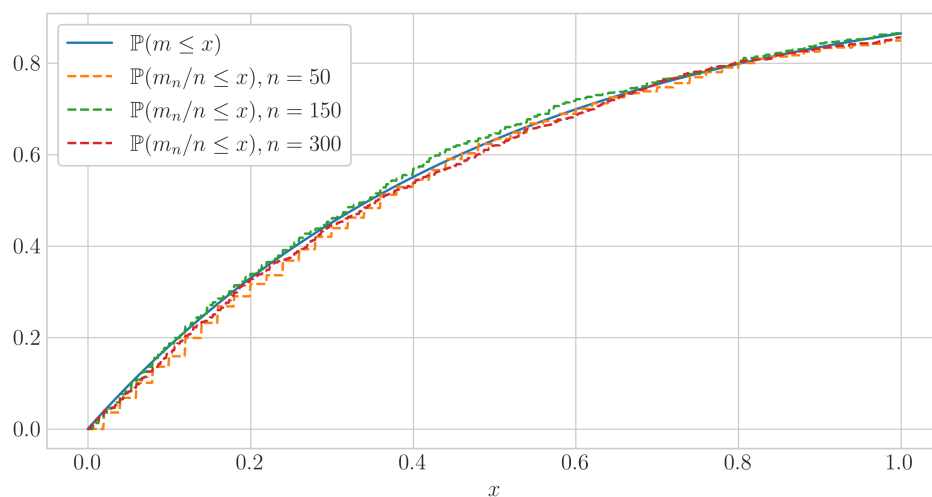


Рисунок 3.6. Функції розподілу  $m$  та  $m_n/n$  для  $n = 50, 150, 300$ .

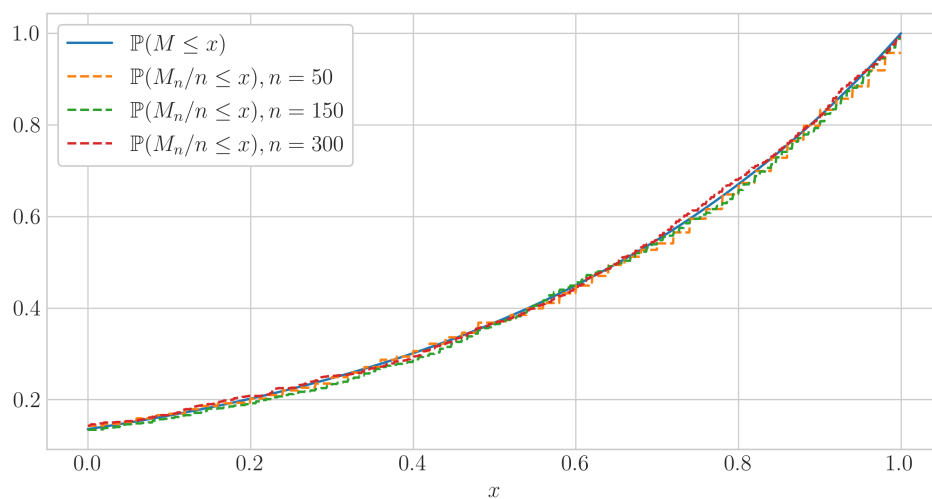


Рисунок 3.7. Функції розподілу  $M$  та  $M_n/n$  для  $n = 50, 150, 300$ .

Для найменшого та найбільшого спейсингів візьмемо  $\theta = 1$ . Знову зауважимо, що, оскільки  $\mathbb{P}(\Psi^{(1)}(\mathbb{X}_1) = 0) = e^{-\theta}$ , то  $\mathbb{P}(\delta = 1) = \mathbb{P}(\Delta = 1) = e^{-\theta}$ , тому у їхніх функцій розподілу є стрибок величиною  $e^{-\theta} \approx 0.367$  в точці 1. Також варто зазначити, що для  $x > 0.5$  маємо  $\mathbb{P}(\delta \leq x) = \mathbb{P}(\Psi^{(1)}(\mathbb{X}_1) > 0) = 1 - e^{-\theta}$ .

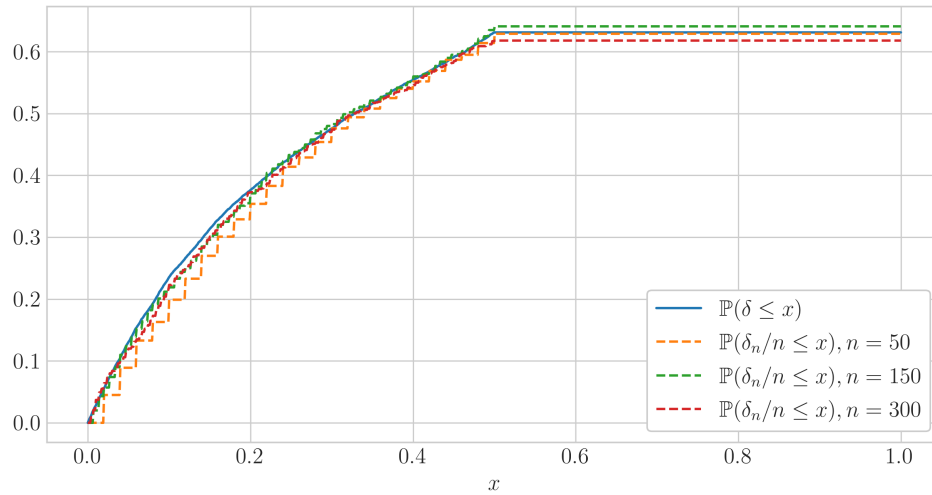


Рисунок 3.8. Функції розподілу  $\delta$  та  $\delta_n/n$  для  $n = 50, 150, 300$ .

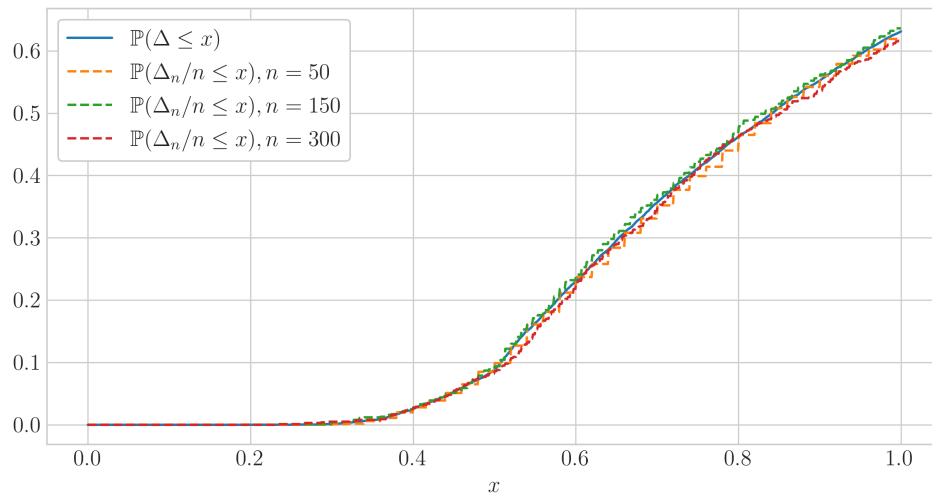


Рисунок 3.9. Функції розподілу  $\Delta$  та  $\Delta_n/n$  для  $n = 50, 150, 300$ .

## ВИСНОВКИ

В даній роботі було розглянуто випадкові перестановки з вагами циклів, задані розподілом (1). Через цикли таких перестановок формулою (3) було визначено послідовність точкових процесів  $\Psi_n$  на спеціально побудованому метричному просторі  $\mathbb{X}$ . В теоремі 3 доведено грубу збіжність за розподілом цієї послідовності до точкового процесу Пуассона  $\Psi$  на  $\mathbb{X}$  з певною мірою інтенсивності.

На кількох прикладах (твердження 1, 2, 3, 4) було показано, як за допомогою властивостей процесу Пуассона та теореми про неперервне відображення (теорема 2) можна отримувати теореми про збіжність за розподілом випадкових величин, пов'язаних з циклами випадкових перестановок. Також було проведено чисельне моделювання отриманих результатів.

Отримані в роботі результати повністю описують асимптотичний склад так званих коротких циклів (циклів обмеженої довжини) випадкових перестановок з розподілом (1). Водночас, ці результати не можна застосувати до так званих довгих циклів (необмеженої довжини) – наприклад, для опису складу найдовшого циклу перестановки. Для цього необхідно шукати інші теоретичні підходи, що може стати темою подальших робіт.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Betz Volker, Ueltschi Daniel. Spatial Random Permutations and Infinite Cycles // *Communications in Mathematical Physics*. 2009. Vol. 285, no. 2. P. 469–501. Access mode: <https://doi.org/10.1007/s00220-008-0584-4>.
2. Betz Volker, Ueltschi Daniel. Spatial Random Permutations and Poisson-Dirichlet Law of Cycle Lengths // *Electronic Journal of Probability*. 2011. Vol. 16, no. none. P. 1173 – 1192. Access mode: <https://doi.org/10.1214/EJP.v16-901>.
3. Betz Volker, Ueltschi Daniel. Spatial random permutations with small cycle weights // *Probability Theory and Related Fields*. 2011. Vol. 149, no. 1. P. 191–222. Access mode: <https://doi.org/10.1007/s00440-009-0248-0>.
4. Billingsley Patrick. Probability and Measure. Wiley Series in Probability & Mathematical Statistics: Probability & Mathematical Statistics. 3 ed. Nashville, TN : John Wiley & Sons, 1995. May.
5. C. Arnold B., N. Balakrishnan, N. Nagaraja H. A First Course in Order Statistics (Classics in Applied Mathematics). USA : Society for Industrial and Applied Mathematics, 2008. 305 p. ISBN: 978-0-89871-906-2.
6. Crane Harry. The Ubiquitous Ewens Sampling Formula // *Statistical Science*. 2016. Vol. 31, no. 1. P. 1–19. Access mode: <http://www.jstor.org/stable/24780825> (online; accessed: 2024-02-23).
7. Dohmen Klaus. Improved Bonferroni Inequalities via Abstract Tubes: Inequalities and Identities of Inclusion-Exclusion Type. Springer Berlin Heidelberg, 2003. ISBN: 9783540393993. Access mode: <http://dx.doi.org/10.1007/b13785>.
8. Ercolani Nicholas M., Ueltschi Daniel. Cycle structure of random permutations with cycle weights // *Random Structures Algorithms*. 2014. Vol. 44, no. 1. P. 109–133.
9. Holst L. On the Lengths of the Pieces of a Stick Broken at Random // *Journal of Applied Probability*. 1980. Vol. 17, no. 3. P. 623–634.
10. Kallenberg O. Random Measures, Theory and Applications. Probability Theory and Stochastic Modelling. Springer International Publishing Switzerland, 2017. 694 p. ISBN: 978-3-319-41598-7.
11. Last G., Penrose M. Lectures on the Poisson Process. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2018. 293 p. ISBN: 978-1-107-08801-6.
12. Montmort Pierre Rémond de. Essay d'analyse sur les jeux de hazard. Paris : J.

Quillau Paris, 1708.

13. NIST handbook of mathematical functions / ed. by Olver Frank W. J., Lozier Daniel W., Boisvert Ronald F., Clark Charles W. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge, 2010. P. xvi+951. ISBN: 978-0-521-14063-8.
14. R. Arratia, A.D. Barbour, S. Tavaré. Poisson process approximations for the Ewens sampling formula // *The Annals of Applied Probability*. 1992. Vol. 2. P. 519–535.
15. R. Arratia, A.D. Barbour, S. Tavaré. Logarithmic combinatorial structures: A probabilistic approach. EMS Monogr. Math. Zürich: European Mathematical Society (EMS), 2003. ISBN: 3-03719-000-0.
16. Resnick S.I. Extreme Values, Regular Variation and Point Processes. Springer Science+Business Media New York, 2008. 320 p. ISBN: 978-0-387-75952-4.
17. S. Tavaré. The magical Ewens sampling formula // *Bulletin of the London Mathematical Society*. 2021. Vol. 53, no. 6. P. 1563–1582.
18. Tables of integral transforms. Vol. I / Erdélyi A., Magnus W., Oberhettinger F., and Tricomi F. G. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1954. P. xx+391.
19. W.J. Ewens. The sampling theory of selectively neutral alleles // *Theoretical Population Biology*. 1972. Vol. 3, no. 1. P. 87–112.
20. Спекторський І. Я. Дискретна математика. Київ, НТУУ «КПІ», НК «ІПСА», 2004. 120 с.

## ДОДАТОК А. ПРОГРАМНИЙ КОД ДЛЯ МОДЕЛЮВАННЯ

```

import numpy as np
from copy import deepcopy
from scipy.stats import poisson, expon
from scipy.special import iv, factorial
from tqdm import tqdm
from joblib import Parallel, delayed
import matplotlib.pyplot as plt
import matplotlib
import os

matplotlib.rcParams["mathtext.fontset"] = "cm"
matplotlib.rcParams["text.usetex"] = True
matplotlib.rcParams["text.latex.preamble"] = r"\usepackage{amsfonts}\renewcommand
{\P}{\mathbb{P}}"
matplotlib.rcParams["font.size"] = "14"
plt.style.use("seaborn-whitegrid")

IMAGE_DIR = "plots_tmp"
RANDOM_STATE = 42
SAMPLE_SIZE = 1000
MARKERS = ["s", "D", "^", "X", "*"]
COLORS = ["tab:blue", "tab:orange", "tab:green", "tab:red"]
DPI_DISPLAY = 80
DPI_SAVE = 300
os.makedirs(IMAGE_DIR, exist_ok=True)

def save_and_show_plot(plt_obj, filename, xlabel="$x$"):
    try: plt_obj[0].xlabel(xlabel)
    except AttributeError: plt_obj[0].set_xlabel(xlabel)
    plt_obj[0].legend(framealpha=1, frameon=True)
    plt_obj[1].savefig(os.path.join(IMAGE_DIR, filename), dpi=DPI_SAVE, bbox_inches=
        "tight")
    plt.show()

def emp_cdf(sample, domain):
    return [(sample <= x).mean() for x in domain]

# Визначення перестановки
class Permutation:
    def __init__(self, cycles):
        self.cycles = deepcopy(cycles)
        self.n = max(max(cycle) for cycle in cycles)
        perm = [0] * self.n
        for cycle in cycles:
            cycle.append(cycle[0])
            for ind in range(len(cycle) - 1):
                perm[cycle[ind] - 1] = cycle[ind + 1]
        if sorted(perm) != list(range(1, self.n + 1)):
            raise ValueError("incorrect set of disjoint cycles")
        self.permutation = perm
        self.cycle_count = len(cycles)

    def __len__(self): return self.n

    def __repr__(self): return repr(self.cycles)

# Процес китайського ресторану
class ChineseRestaurantProcess:
    def __init__(self, theta, n):
        assert theta > 0, "theta must be > 0"
        self.theta = theta
        self.n = n

    def _sample_A(self, i, rng):
        p = [1 / (self.theta + i - 1) for j in range(i)]

```



```

p[-1] = p[-1] * self.theta
A = rng.choice(range(1, i + 1), p=p)
return A

def _sample_cycles(self, random_state):
    rng = np.random.default_rng(random_state)
    cycles = [[1]]
    for i in range(2, self.n + 1):
        A = self._sample_A(i, rng)
        if A == i: cycles.append([i])
        else:
            for cycle_ind, cycle in enumerate(cycles):
                if A in cycle:
                    ind = cycle.index(A)
                    break
            cycles[cycle_ind].insert(ind + 1, i)
    return cycles

def _sample_one_permutation(self, random_state=None):
    cycles = self._sample_cycles(random_state=random_state)
    return Permutation(cycles)

def _sample_cycles_fixed_len(self, cycle_len, random_state=None):
    all_cycles = self._sample_cycles(random_state=random_state)
    return list(filter(lambda c: len(c) == cycle_len, all_cycles))

def permutations(self, n_samples, verbose=False, random_state=None):
    rng = np.random.default_rng(random_state)
    generator = tqdm(range(n_samples)) if verbose else range(n_samples)
    sample = Parallel(n_jobs=-1)(
        delayed(self._sample_one_permutation)(random_state=rng.integers(n_samples
**2))
        for _ in generator)
    return sample

def cycles(self, n_samples, cycle_len, verbose=False, random_state=None):
    rng = np.random.default_rng(random_state)
    generator = tqdm(range(n_samples)) if verbose else range(n_samples)
    sample = Parallel(n_jobs=-1)(
        delayed(self._sample_cycles_fixed_len)(
            cycle_len=cycle_len, random_state=rng.integers(n_samples**2)) for _ in
generator)
    return sample

# Розподіл кількості циклів
def CycleCountsSample(n, theta, bounds, sample_size=1, random_state=None,
    aggregate=False, verbose=False):
    def _check_bound(n, cycle, bound):
        return all(interval[0] < i / n < interval[1] for i, interval in zip(cycle,
bound))

    def _count_within_bounds(n, permutation, bounds, aggregate):
        res = {key: 0 for key in bounds.keys()}
        for cycle in permutation.cycles:
            k = len(cycle)
            if k in bounds:
                res[k] += any(_check_bound(n, cycle, bound) for bound in bounds[k])
        if aggregate: return sum(res.values())
        else: return res

    sampler = ChineseRestaurantProcess(theta, n)
    if verbose:
        print(f"CycleCountsSample, n = {n}, theta = {theta}, bounds = {bounds},
sample_size = {sample_size}")
    permutations_sample = sampler.permutations(sample_size, verbose=verbose,
random_state=random_state)

```

```

counts = Parallel(n_jobs=-1)(
    delayed(lambda x: _count_within_bounds(n, x, bounds, aggregate))(permutation)
    for permutation in permutations_sample)
if aggregate: res = counts
else:
    res = {key: [] for key in bounds}
    for count in counts:
        for key in res:
            res[key].append(count[key])
return res

def measure(bound):
    res = 1
    for interval in bound: res *= interval[1] - interval[0]
    return res

n_range = [100, 250, 500]
theta = 10
bounds = {
    1: [(0.65, 0.85),],
    2: [(0.25, 0.4), (0.5, 0.7)], [(0.05, 0.1), (0.15, 0.2)],
    3: [(0.2, 0.3), (0.9, 1), (0.35, 0.45)]
}
bounds_measures = {key: sum(measure(bound) for bound in bounds[key]) for key in
    bounds}
set_measure = sum(bounds_measures.values())
print(f"measure: {set_measure:.5f}")
plt.figure(figsize=(10, 5), dpi=DPI_DISPLAY)
X = poisson(mu=theta * set_measure)
plt.plot(range(12), X.pmf(range(12)), marker="o", markersize=5, label=r"$\P(X = k)$",
    c=COLORS[0])
for n, marker, color in zip(n_range, MARKERS, COLORS[1:]):
    res = CycleCountsSample(n, theta, bounds, 5 * SAMPLE_SIZE, RANDOM_STATE,
        aggregate=True, verbose=True)
    values, counts = np.unique(res, return_counts=True)
    freqs = counts / sum(counts)
    plt.plot(values, freqs, marker=marker, markersize=5,
        label=rf"$\P(X_n = k), n = {n}$", c=color, linestyle="--")
save_and_show_plot((plt, plt), f"cycle_counts_theta_{theta}.png", xlabel="$k$")

# Розмахи циклів
def CycleRangesSample(n, theta, cycle_len, sample_size=1, random_state=None,
    verbose=False):
    sampler = ChineseRestaurantProcess(theta, n)
    if verbose:
        print(f"CycleRangesSample, n = {n}, theta = {theta}, cycle_len = {cycle_len},
            sample_size = {sample_size}")
    cycles_sample = sampler.cycles(sample_size, cycle_len, verbose, random_state)
    cycle_ranges = Parallel(n_jobs=-1)(delayed(lambda x: [max(e) - min(e) for e in x
    ])(cycle)
    for cycle in cycles_sample)
    return cycle_ranges

def MinCycleRange(cycle_ranges_sample, n):
    min_cycle_range = [min(r) if len(r) else n for r in cycle_ranges_sample]
    return np.array(min_cycle_range) / n

def MaxCycleRange(cycle_ranges_sample, n):
    max_cycle_range = [max(r) if len(r) else 0 for r in cycle_ranges_sample]
    return np.array(max_cycle_range) / n

def MinCycleRangeCDF(x, theta, cycle_len):
    k = cycle_len
    p = k * (x ** (k - 1)) - (k - 1) * (x**k)
    return np.clip(1 - np.exp(-theta * p / k), 0, 1)

```

```

def MaxCycleRangeCDF(x, theta, cycle_len):
    k = cycle_len
    p = k * (x ** (k - 1)) - (k - 1) * (x**k) - 1
    return np.clip(np.exp(theta * p / k), 0, 1)

n_range = [100, 250, 500]
cycle_len = 4
theta = 3
fig_min, ax_min = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
fig_max, ax_max = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
domain = np.linspace(0, 0.9999, 1000)
ax_min.plot(domain, MinCycleRangeCDF(domain, theta, cycle_len), label=r"$\P(r^{\{k\}} \le x)$")
ax_max.plot(domain, MaxCycleRangeCDF(domain, theta, cycle_len), label=r"$\P(R^{\{k\}} \le x)$")
for n in n_range:
    cycle_ranges = CycleRangesSample(n, theta, cycle_len, SAMPLE_SIZE, RANDOM_STATE,
    verbose=True)
    min_cycle_range = MinCycleRange(cycle_ranges, n)
    max_cycle_range = MaxCycleRange(cycle_ranges, n)
    ax_min.plot(domain, emp_cdf(min_cycle_range, domain),
    label=rf"$\P(r_n^{\{k\}} \hspace{-0.2em}/n \le x), n = \{n\}$", linestyle="--")
    ax_max.plot(domain, emp_cdf(max_cycle_range, domain),
    label=rf"$\P(R_n^{\{k\}} \hspace{-0.2em}/n \le x), n = \{n\}$", linestyle="--")
    save_and_show_plot((ax_min, fig_min), f"min_cycle_range_theta_{theta}_k_{cycle_len}.png")
    save_and_show_plot((ax_max, fig_max), f"max_cycle_range_theta_{theta}_k_{cycle_len}.png")

# Статистики нерухомих точок
def FixedPointsSample(n, theta, sample_size=1, random_state=None, verbose=False):
    sampler = ChineseRestaurantProcess(theta, n)
    if verbose:
        print(f"FixedPointsSample, n = {n}, theta = {theta}, sample_size = {sample_size}")
    fp_sample = sampler.cycles(sample_size, 1, verbose, random_state)
    fp_sample_flat = Parallel(n_jobs=-1)(delayed(lambda x: sum(x, []))(fps) for fps in fp_sample)
    return fp_sample_flat

## Сума нерухомих точок
def FixedPointsSum(fixed_points_sample, n):
    max_cycle_range = [sum(fps) if len(fps) else 0 for fps in fixed_points_sample]
    return np.array(max_cycle_range) / n

def FixedPointsSumCDF(x, theta):
    if not hasattr(x, "__iter__"):
        dom = [x]
    dom = np.array(x).copy()
    dom[dom < 0] = 0
    n = int(np.floor(dom.max()))
    min_x = int(np.floor(dom.min()))
    ar = np.arange(min_x, n + 1)
    k = np.tile(np.arange(n + 1), (n + 1, 1))
    k[:min_x] = -1
    k = k[np.tril_indices(n + 1)]
    k = np.expand_dims(k[k >= 0], axis=1)
    mask = (dom >= ar.reshape(len(ar), 1)) * (dom < (ar.reshape(len(ar), 1) + 1))
    dom = np.repeat(dom * mask, ar + 1, axis=0)
    dom = dom - k
    dom = np.sqrt(theta * dom * (dom >= 0))
    res = (dom != 0) * np.power(-dom, k) * iv(k, 2 * dom) / factorial(k)
    res = np.exp(-theta) * res.sum(axis=0)
    res[x == 0] = np.exp(-theta)

```

```

if not hasattr(x, "__iter__"): return res[0]
else: return res

n_range = [50, 150, 300]
theta = 5
plt.figure(figsize=(10, 5), dpi=DPI_DISPLAY)
domain = np.linspace(0, 2 * theta, 1000)
plt.plot(domain, FixedPointsSumCDF(domain, theta), label=r"$\P(S^{\{1\}} \le x)$")
for n in n_range:
    fixed_points_sample = FixedPointsSample(n, theta, SAMPLE_SIZE, RANDOM_STATE,
        verbose=True)
    fixed_points_sum = FixedPointsSum(fixed_points_sample, n)
    plt.plot(domain, emp_cdf(fixed_points_sum, domain),
        label=rf"$\P(S_n^{\{1\}} \le x), n = {n}$", linestyle="--")
save_and_show_plot((plt, plt), f"fixed_points_sum_theta_{theta}.png")

## Найменша та найбільша нерухомі точки
def MinFixedPointSample(fixed_points_sample, n):
    min_fixed_point = [min(r) if len(r) else n for r in fixed_points_sample]
    return np.array(min_fixed_point) / n

def MaxFixedPointSample(fixed_points_sample, n):
    max_fixed_point = [max(r) if len(r) else 0 for r in fixed_points_sample]
    return np.array(max_fixed_point) / n

def MinFixedPointCDF(x, theta):
    return np.clip(1 - np.exp(-theta * x), 0, 1)

def MaxFixedPointCDF(x, theta):
    return np.clip(np.exp(theta * (x - 1)), 0, 1)

n_range = [50, 150, 300]
theta = 2
fig_min, ax_min = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
fig_max, ax_max = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
domain = np.linspace(0, 0.9999, 1000)
ax_min.plot(domain, MinFixedPointCDF(domain, theta), label=r"$\P(m \le x)$")
ax_max.plot(domain, MaxFixedPointCDF(domain, theta), label=r"$\P(M \le x)$")
for n in n_range:
    fixed_points_sample = FixedPointsSample(n, theta, SAMPLE_SIZE, RANDOM_STATE,
        verbose=True)
    min_fixed_point = MinFixedPointSample(fixed_points_sample, n)
    max_fixed_point = MaxFixedPointSample(fixed_points_sample, n)
    ax_min.plot(domain, emp_cdf(min_fixed_point, domain),
        label=rf"$\P(m_n/n \le x), n = {n}$", linestyle="--")
    ax_max.plot(domain, emp_cdf(max_fixed_point, domain),
        label=rf"$\P(M_n/n \le x), n = {n}$", linestyle="--")
save_and_show_plot((ax_min, fig_min), f"min_fixed_point_theta_{theta}.png")
save_and_show_plot((ax_max, fig_max), f"max_fixed_point_theta_{theta}.png")

## Спейсинги
def MinSpacing(fixed_points_sample, n):
    fixed_points = np.zeros((len(fixed_points_sample), n))
    for ind, fps in enumerate(fixed_points_sample):
        fps_np = np.array(fps)
        if len(fps_np): fixed_points[ind][fps_np - 1] = fps_np
    spacings = np.diff(np.maximum.accumulate(fixed_points, axis=1), axis=1)
    fixed_points_min = fixed_points.copy()
    fixed_points_min[fixed_points_min == 0] = n + 1
    spacings = np.c_[fixed_points_min.min(axis=1), spacings, n - fixed_points.max(
        axis=1) + 1]
    spacings[spacings == 0] = n + 1

```

```

min_spacings = spacings.min(axis=1)
min_spacings[min_spacings == n + 1] = n
return min_spacings / n

def MaxSpacing(fixed_points_sample, n):
    fixed_points = np.zeros((len(fixed_points_sample), n))
    for ind, fps in enumerate(fixed_points_sample):
        fps_np = np.array(fps)
        if len(fps_np): fixed_points[ind][fps_np - 1] = fps_np
    spacings = np.diff(np.maximum.accumulate(fixed_points, axis=1), axis=1)
    fixed_points_min = fixed_points.copy()
    fixed_points_min[fixed_points_min == 0] = n + 1
    spacings = np.c_[fixed_points_min.min(axis=1), spacings, n - fixed_points.max(
        axis=1) + 1]
    max_spacings = spacings.max(axis=1)
    max_spacings[max_spacings == n + 1] = n
    return max_spacings / n

def sample_delta(theta, sample_size, random_state=None):
    rng = np.random.default_rng(random_state)
    nu = poisson(mu=theta).rvs(size=sample_size, random_state=random_state)
    res = []
    for n in nu:
        X = expon.rvs(size=n + 1, random_state=rng.integers(sample_size**2))
        res.append(X[n] / ((n + 1) * X.sum()))
    return np.array(res)

def sample_Delta(theta, sample_size, random_state=None):
    rng = np.random.default_rng(random_state)
    nu = poisson(mu=theta).rvs(size=sample_size, random_state=random_state)
    res = []
    for n in nu:
        X = expon.rvs(size=n + 1, random_state=rng.integers(sample_size**2))
        res.append((X / np.arange(1, n + 2)).sum() / X.sum())
    return np.array(res)

n_range = [50, 150, 300]
theta = 1
fig_min, ax_min = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
fig_max, ax_max = plt.subplots(figsize=(10, 5), dpi=DPI_DISPLAY)
delta = sample_delta(theta, 10 * SAMPLE_SIZE, RANDOM_STATE)
Delta = sample_Delta(theta, 10 * SAMPLE_SIZE, RANDOM_STATE)
domain = np.linspace(0, 0.9999, 1000)
ax_min.plot(domain, emp_cdf(delta, domain), label=r"$P(\delta \le x)$")
ax_max.plot(domain, emp_cdf(Delta, domain), label=r"$P(\Delta \le x)$")
for n in n_range:
    fixed_points_sample = FixedPointsSample(n, theta, SAMPLE_SIZE, RANDOM_STATE,
        verbose=True)
    min_spacing = MinSpacing(fixed_points_sample, n)
    max_spacing = MaxSpacing(fixed_points_sample, n)
    ax_min.plot(domain, emp_cdf(min_spacing, domain),
        label=rf"$P(\delta_n/n \le x), n = {n}$", linestyle="--")
    ax_max.plot(domain, emp_cdf(max_spacing, domain),
        label=rf"$P(\Delta_n/n \le x), n = {n}$", linestyle="--")
save_and_show_plot((ax_min, fig_min), f"min_spacing_theta_{theta}.png")
save_and_show_plot((ax_max, fig_max), f"max_spacing_theta_{theta}.png")

```