

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

*Б.В. Довгай*

# Математичні основи генетики популяцій

Електронний навчальний посібник

Електронна бібліотека факультету кібернетики

Київ

2014

# ЗМІСТ

1.	Генетичні моделі. Опис генетичного механізму . . . . .	4
1.1.	Розмноження і закон Харді-Вайнберга . . . . .	4
1.2.	Мутація . . . . .	7
1.3.	Міграція . . . . .	8
1.4.	Відбір . . . . .	8
1.5.	Моделі двостатевих популяцій . . . . .	11
2.	Ланцюги Маркова . . . . .	14
2.1.	Означення . . . . .	14
2.2.	Матриці ймовірностей переходу ланцюга Маркова .	16
2.3.	Класифікація станів ланцюга Маркова . . . . .	17
2.4.	Періодичність ланцюга Маркова . . . . .	17
2.5.	Рекурентність . . . . .	18
2.6.	Основні граничні теореми для ланцюгів Маркова .	19
2.7.	Ймовірності поглинання . . . . .	21
3.	Інбридинг . . . . .	24
3.1.	Самозапліднення . . . . .	24
3.2.	Спарювання сібсів . . . . .	25
3.3.	Гени, ідентичні за походженням . . . . .	28
3.4.	Нескінченна диплоїдна популяція та коефіцієнт інбридингу . . . . .	30
3.5.	Популяція з самозаплідненням та випадковим спарюванням . . . . .	31
4.	Поліплоїди . . . . .	32
5.	Гіллясті процеси . . . . .	35
5.1.	Означення . . . . .	36
5.2.	Співвідношення для генератрис, що описують гіллястий процес . . . . .	37
5.3.	Ймовірності виродження . . . . .	41

5.4.	Приклади . . . . .	43
6.	Процеси Маркова, що породжуються гіллястими процесами	47
6.1.	Модель Райта . . . . .	47
6.2.	Модель полісомного спадкування . . . . .	49
6.3.	Процеси Маркова, що породжуються прямим до- бутком гіллястих процесів . . . . .	50

# 1. Генетичні моделі. Опис генетичного механізму

В ядрі клітини знаходяться *хромосоми*, які відповідають за спадкові характеристики організму. Їх кількість різна для різних організмів. У більшості видів організмів кількість хромосом в клітині коливається від 10 до 50. В клітинах людини та багатьох тварин по 46 хромосом.

Одиниця, що контролює повністю або частково спадкування деякої ознаки, називається *геном*. Місце, яке ген займає в хромосомі, називається *локусом*.

Набір хромосом може містити по одній хромосомі кожного типу (*гаплоїди*), як у деяких грибів, по парі хромосом (*диплоїди*), як у ссавців, або групи більшого розміру (*триплоїди*, *тетраплоїди* та в загальному випадку *поліплоїди*), як у багатьох рослин.

В кожному локусі гени можуть належати до різних контрастуючих типів, які називаються *алелями*. Так, деякий ген, що відповідає, наприклад, за карий колір очей, є алелем гена, що відповідає за блакитний колір очей. Різні можливі комбінації алелей у відповідних локусах різних хромосом одного типу називаються *генотипами*. При диплоїдії, якщо є алелі  $A$  та  $a$ , можливі наступні генотипи:  $AA$ ,  $Aa$ ,  $aa$ . Генотипи  $AA$  і  $aa$  називаються *гомозиготними*, а генотип  $Aa$  — *гетерозиготним*.

*Фенотипами* називаються різноманітні зовнішні прояви алельних комбінацій. Наприклад, якщо  $A$  домінує над  $a$ , то різні генотипи  $AA$  та  $Aa$  будуть визначати один і той же фенотип, а комбінація  $aa$  визначає інший фенотип.

## 1.1. Розмноження і закон Харді-Вайнберга

Існує два способи клітинного ділення. Більш поширеним є *мітоз*. При цьому способі кожна хромосома ділиться на дві і, коли клітина розщеплюється на дві дочірні клітини, кожна з новостворених клітин отримує по одній хромосомі з кожної пари. Таким чином, кожна нова клітина має повний набір хромосом.

Процес другого типу — *мейоз* — відбувається при статевому розмно-

женні диплоїдних клітин. Кожна хромосома при цьому подвоюється. Потім утворюються чотири нові клітини, кожна з яких містить одну хромосому. Таким чином, кожна з чотирьох нових клітин містить половину нормального набору хромосом. Така клітина називається *гаметою*. Коли дві гамети з'єднуються, утворюється запліднена клітина — *зигота*, яка має повний набір хромосом.

Розглянемо тепер якийсь локус в парі хромосом у диплоїдного організму для випадку, коли можливі два алелі  $A$  та  $a$ . Надалі будемо розглядати нескінченну популяцію, різні покоління якої не змішуються.

Припустимо, що схрещується генотип  $AA$  з генотипом  $Aa$ . Перший з них утворює лише гамети  $A$ , другий — гамети  $A$  і  $a$  в рівних пропорціях. При спарюванні (схрещуванні), що розглядається, гамети, утворені генотипом  $AA$ , очевидно, однаково часто будуть з'єднуватись з різними гаметами, утвореними генотипом  $Aa$ . Таким чином, для випадку спарювання  $AA \times Aa$  частота нащадків типу  $AA$  дорівнює  $\frac{1}{2}$ , і така ж частота для нащадків типу  $Aa$ . Аналогічно при спарюванні  $AA \times aa$  утворюється тільки нащадки типу  $Aa$ . Спарювання  $Aa \times Aa$  дає наступне співвідношення нащадків:  $\frac{1}{4}AA$ ,  $\frac{1}{2}Aa$ ,  $\frac{1}{4}aa$  і т.д. Ці пропорції відомі під назвою *співвідношення при розщепленні*.

Розглянемо деяку роздільностатеву популяцію з трьома генотипами  $AA$ ,  $Aa$ ,  $aa$ , в якій відбувається випадкове спарювання. Нехай вихідне співвідношення генотипів наступне: частка  $AA$  дорівнює  $u$ , частка  $Aa$  дорівнює  $v$  і частка  $aa$  дорівнює  $w$ ,  $u + v + w = 1$ . Складемо таблицю результатів випадкового спарювання:

Схрещування	Типи нащадків	Частота даного типу схрещування
$AA \times AA$	$AA$	$u^2$
$AA \times Aa$	$\frac{1}{2}AA + \frac{1}{2}Aa$	$2uv$
$AA \times aa$	$Aa$	$2uw$
$Aa \times Aa$	$\frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$	$v^2$
$Aa \times aa$	$\frac{1}{2}Aa + \frac{1}{2}aa$	$2vw$
$aa \times aa$	$aa$	$w^2$

Отже, в наступному поколінні ці три генотипи будуть представлені в співвідношенні:

$$\begin{array}{ccc}
 AA & Aa & aa \\
 u^2 + uv + \frac{1}{4}v^2 & 2uw + vw + uv + \frac{1}{2}v^2 & w^2 + vw + \frac{1}{4}v^2
 \end{array}$$

Покладемо  $p = u + \frac{v}{2}$ ,  $q = w + \frac{v}{2}$ , тоді ці пропорції набудуть вигляду

$$\begin{array}{ccc}
 AA & Aa & aa \\
 p^2 & 2pq & q^2
 \end{array}$$

і  $p + q = 1$ .

Точно таким же чином ми знайдемо, що після наступного спарювання генотипи розподіляться у відношенні

$$\begin{array}{ccc}
 AA & Aa & aa \\
 (p^2 + pq)^2 = p^2 & 2(p^2 + pq)(q^2 + pq) = 2pq & (q^2 + pq)^2 = q^2
 \end{array}$$

Таким чином, починаючи з другого покоління, співвідношення між генотипами не змінюється, тобто стійкий стан досягається за одне покоління. Цей результат відомий, як *закон Харді-Вайнберга*.

Можна отримати цей же результат наступним чином. Розглянемо популяцію гамет: якщо початкове співвідношення між генотипами те саме, що й раніше, то відповідна частка гамет  $A$  в популяції дорівнює  $u + \frac{v}{2} = p$ , а гамет  $a$  дорівнює  $w + \frac{v}{2} = q$ , отже в наступному поколінні генотипи представлені у відношенні

$$\begin{array}{ccc} AA & Aa & aa \\ p \cdot p = p^2 & 2pq & q \cdot q = q^2 \end{array}$$

Розглядаючи сукупність гамет в новій популяції, знайдемо:  
 частка ( $A$ ) =  $p^2 + pq = p(p + q) = p$ ,  
 частка ( $a$ ) =  $q^2 + pq = q(p + q) = q$ ,  
 що вказує на досягнену стійкість.

Факт співпадіння результатів, отриманих обома методами, відомий у вигляді наступного твердження: “випадкове злиття гамет” еквівалентне випадковому спарюванню.

Тепер ми розглянемо, як три фактори — мутація, міграція та відбір — впливають на популяцію генотипів у припущенні про випадкове спарювання

## 1.2. Мутація

1. Розглянемо гапліодну популяцію генів  $A$  та  $a$ . Припустимо, що в проміжку між будь-якими двома наступними поколіннями частка  $\mu$  генів  $a$  мутує (змінюється) в форму  $A$ . Якщо  $p_n$  — частка генів  $a$  в  $n$ -му поколінні, то  $(1 - p_n)$  — частка генів  $A$  в  $n$ -му поколінні. Тоді  $p_{n+1} = p_n(1 - \mu) = p_{n-1}(1 - \mu)^2 \dots$  і остаточно  $p_n = p_0(1 - \mu)^n$ , тобто, якщо  $0 < \mu < 1$ , то  $p_n \rightarrow 0$ ,  $n \rightarrow \infty$  по геометричній прогресії із знаменником  $1 - \mu$ .

2. Нехай тепер між двома наступними поколіннями разом з розглянутою мутацією частка  $\nu > 0$  генів  $A$  мутує в форму  $a$ . Тоді

$$p_{n+1} = p_n(1 - \mu) + \nu(1 - p_n) = p_n(1 - \mu - \nu) + \nu,$$

що еквівалентно співвідношенню

$$p_{n+1} - \frac{\nu}{\mu + \nu} = \left( p_n - \frac{\nu}{\mu + \nu} \right) (1 - \mu - \nu),$$

бо

$$\nu = (\mu + \nu) \frac{\nu}{\mu + \nu} = (1 - (1 - \mu - \nu)) \frac{\nu}{\mu + \nu} = \frac{\nu}{\mu + \nu} - \frac{\nu}{\mu + \nu} (1 - \mu - \nu).$$

За індукцією отримаємо

$$p_{n+1} - \frac{\nu}{\mu + \nu} = \left( p_{n-1} - \frac{\nu}{\mu + \nu} \right) (1 - \mu - \nu)^2 = \dots = \left( p_0 - \frac{\nu}{\mu + \nu} \right) (1 - \mu - \nu)^n.$$

Знову збіжність має геометричний характер. Зауважимо, що якщо  $\mu + \nu > 1$ ,  $1 - \mu - \nu < 0$ , то  $p_n$  буде збігатися до границі немонотонно, якщо ж  $\mu + \nu \leq 1$ , то збіжність монотонна. Помітимо, що оскільки  $\mu, \nu < 1$ , то в будь-якому випадку  $|1 - \mu - \nu| < 1$ , а отже, збіжність дійсно має місце:  $p_n \xrightarrow{n \rightarrow \infty} \frac{\nu}{\mu + \nu}$ .

### 1.3. Міграція

Припустимо, що ми маємо 2 популяції, що складаються з генів  $A$  та  $a$ , — внутрішню популяцію, в якій частоти цих генів дорівнюють  $p_n$  та  $q_n$  відповідно (і їх ми хочемо знайти), і зовнішню “велику” популяцію, в якій гени  $A$  та  $a$  мають постійні частоти  $P$  та  $Q$ . “Міграція” означає, що частка  $I$  усієї внутрішньої популяції за час одного покоління заміщується такою ж кількістю індивідумів, взятих із зовнішньої популяції. Тоді

$$p_{n+1} = p_n(1 - I) + PI.$$

Оскільки величина  $P$  стала, отримане рекурентне співвідношення співпадає, очевидно, з тим, що описує процес мутації, якщо взяти  $\nu = PI$ ,  $\mu = (1 - P)I$ , бо  $\mu = I - \nu$ ,  $1 - \mu - \nu = 1 - I$ . Звідси випливає, що  $p_n$  збігається до величини  $\frac{\nu}{\mu + \nu} = \frac{PI}{I - PI + PI} = P$ .

### 1.4. Відбір

Існує 2 види відбору — *гаметичний*, який відображає відносні переваги деяких гамет над іншими, та *зиготний*, який відображає переваги деяких генотипів над іншими. Такі переваги можуть відповідати різним смертностям, відносній плодючості і т.д.

#### (а) Гаметичний відбір

Нехай  $1$  і  $1 - s$  ( $0 < s < 1$ ) — міри відносної селективної переваги гамет  $A$  в порівнянні з гаметою  $a$ . Нехай  $p_n$  і  $q_n$  означають їх частоти



в популяції в  $n$ -му поколінні. Ці дані можна звести в таблицю

	Гамета $A$	Гамета $a$	
Частота	$p_n$	$q_n$	$p_n + q_n = 1$
Відносна перевага	1	$1 - s$	

Сформує наступне покоління гамет  $A$  і  $a$ , припускаючи наявним частотам гамет ваги 1 і  $1 - s$  відповідно. Отримаємо

$$p_{n+1} = p_n [p_n + (1 - s)q_n]^{-1}; \quad q_{n+1} = (1 - s)q_n [p_n + (1 - s)q_n]^{-1}.$$

Співмножник  $[p_n + (1 - s)q_n]^{-1}$  є нормуючим.

Маємо

$$p_{n+1} = \frac{p_n}{1 - q_n s}; \quad q_{n+1} = \frac{(1 - s)q_n}{1 - q_n s}.$$

Звідси

$$\frac{q_{n+1}}{p_{n+1}} = (1 - s) \frac{q_n}{p_n} = \dots = (1 - s)^{n+1} \frac{q_0}{p_0} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Отже,  $q_n \rightarrow 0$  і  $p_n \rightarrow 1$ , тобто популяція генів  $a$  вироджується.

### (б) Зиготний відбір. Загальний випадок.

Розглянемо диплоїдну популяцію, в якій кожен генотип характеризується індексом  $\sigma$ , який відображає його перевагу або недолік у порівнянні з іншими типами. Припустимо, що відносна перевага не залежить від часу та частот гамет в  $n$ -му поколінні.

	$AA$	$Aa$	$aa$
Частота в $n$ -му поколінні	$p_n^2$	$2p_n q_n$	$q_n^2$
Відносна перевага	$\sigma_1$	$\sigma_2$	$\sigma_3$

Нехай  $p_n$  і  $q_n$  — частота  $A$  і  $a$  відповідно в  $n$ -му поколінні.

Зиготні частоти  $AA$ ,  $Aa$  і  $aa$  в  $(n + 1)$ -му поколінні дорівнюють відповідно

$$\frac{\sigma_1 p_n^2}{w}, \quad \frac{\sigma_2 2p_n q_n}{w}, \quad \frac{\sigma_3 q_n^2}{w},$$

де величина  $w = \sigma_1 p_n^2 + 2\sigma_2 p_n q_n + \sigma_3 q_n^2$  називається *функцією пристосованості* і  $w^{-1}$  грає роль нормуючого множника. Отже, частоти гамет

А і а дорівнюють відповідно

$$p_{n+1} = \frac{\sigma_1 p_n^2 + \sigma_2 p_n q_n}{w}, \quad q_{n+1} = \frac{\sigma_3 q_n^2 + \sigma_2 p_n q_n}{w}. \quad (1.1)$$

Таким чином,

$$\begin{aligned} \Delta p_n &= p_{n+1} - p_n = \frac{\sigma_1 p_n^2 + \sigma_2 p_n q_n}{w} - p_n \\ &= \frac{\sigma_1 p_n^2 + \sigma_2 p_n q_n - \sigma_1 p_n^3 - 2\sigma_2 p_n^2 q_n - \sigma_3 p_n q_n^2}{w} \\ &= \frac{\sigma_1 p_n^2 + \sigma_2 p_n q_n (q_n - p_n) - \sigma_3 p_n q_n^2}{w} \\ &= p_n q_n \frac{\sigma_1 p_n + \sigma_2 q_n - \sigma_2 p_n - \sigma_3 q_n}{w} = p_n q_n \frac{(\sigma_1 - \sigma_2)p_n + (\sigma_2 - \sigma_3)q_n}{w}. \end{aligned}$$

Тобто

$$\Delta p_n = p_n q_n \frac{(\sigma_1 - \sigma_2)p_n + (\sigma_2 - \sigma_3)q_n}{w}. \quad (1.2)$$

Рівноважною частотою є будь-яка величина  $\tilde{p}$  ( $0 \leq \tilde{p} \leq 1$ ) така, що з рівності  $p_n = \tilde{p}$  випливає  $p_{n+1} = \tilde{p}$ . Вона називається стійкою, якщо для значень  $p_n$ , близьких до  $\tilde{p}$ , змінна  $p_{n+k}$  при  $k \rightarrow \infty$  прямує до  $\tilde{p}$ . Рівноважна частота, яка не є стійкою, називається нестійкою. Визначимо *поліморфізм* як такий стійкий рівноважний стан, в якому представлений кожен генотип. Така ситуація має місце, коли гетерозиготний індивідуум має селективну перевагу перед кожним гомозиготним індивідуумом, як, наприклад, при спадкуванні ознаки серповидноклітинної анемії.

**Приклад поліморфізма:**  $\sigma_2 > \sigma_1$ ,  $\sigma_2 > \sigma_3$ . Така ситуація зазвичай називається гетерозиготною перевагою, або гетерозиготним домінуванням. Розглянемо випадки, коли  $\Delta p_n$  дорівнює нулю.

- (а)  $p_n = 1$ . Це рівноважний стан. Проте, якщо  $p_n$  близьке до 1, то  $\Delta p_n < 0$  і змінна  $p_n$  віддаляється від 1. Це нестійка рівновага.
- (б)  $p_n = 0$ . Цей стан також нестійкий, оскільки  $\Delta p_n > 0$  при малих  $p_n$ .

$$(B) \frac{p_n}{q_n} = \frac{\sigma_2 - \sigma_3}{\sigma_2 - \sigma_1}, \text{ або, що те саме, } p_n = p^*, \text{ де}$$

$$p^* = \frac{\sigma_2 - \sigma_3}{2\sigma_2 - \sigma_1 - \sigma_3}, \quad 0 < p^* < 1.$$

Нехай  $p = p^* + \delta$ . Тоді

$$\begin{aligned} \Delta p &= \frac{pq}{w(p)} ((\sigma_1 - \sigma_2)(p^* + \delta) + (\sigma_2 - \sigma_3)(1 - p^* - \delta)) \\ &= \frac{pq}{w(p)} ((\sigma_1 - \sigma_2)p^* + (\sigma_1 - \sigma_2)\delta + (\sigma_2 - \sigma_1)p^* - (\sigma_2 - \sigma_3)\delta) \\ &= \frac{(p^* + \delta)(q^* - \delta)}{w(p^* + \delta)} (\sigma_1 + \sigma_3 - 2\sigma_2)\delta. \end{aligned}$$

Отже,  $\Delta p \simeq \frac{p^* q^*}{w(p^*)} (\sigma_1 + \sigma_3 - 2\sigma_2)\delta$  має знак, протилежний до знаку  $\delta$ , оскільки  $\sigma_1 + \sigma_3 - 2\sigma_2 < 0$ . Звідси випливає, що стан рівноваги стійкий.

Трохи більш детальний аналіз показує, що процес (1.1) збігається до рівноважного стану  $p^*$  при будь-якому початковому стані  $p$ , такому, що  $p(1-p) \neq 0$ .

## 1.5. Моделі двостатевих популяцій

Розглянемо тепер популяцію, що складається з чоловічих та жіночих особин, спарювання при цьому відбувається лише між представниками різних статей. Якщо обмежитися розглядом лише генів, що локалізовані на аутосомних (нестатевих) хромосомах (як це було в попередньому випадку), то процес випадкового спарювання можна описати таким же чином, як і раніше. Припустимо, що частоти генів всередині чоловічої та жіночої популяцій дорівнюють:

Чоловіча популяція			Жіноча популяція		
$AA$	$Aa$	$aa$	$AA$	$Aa$	$aa$
$r_1$	$2s_1$	$t_1$	$u_1$	$2v_1$	$w_1$

Гамети мають наступні частоти:

Чоловіча популяція		Жіноча популяція	
A	a	A	a
$p_1 = r_1 + s_1$	$q_1 = s_1 + t_1$	$\bar{p}_1 = u_1 + v_1$	$\bar{q}_1 = v_1 + w_1$

Для випадкового спарювання, яке еквівалентне випадковому вибору з чоловічих та жіночих сукупностей гамет, ми знайдемо, що частоти генотипів в обох чоловічій та жіночій популяціях у другому поколінні дорівнюють  $p_1\bar{p}_1$ ,  $p_1\bar{q}_1 + \bar{p}_1q_1$ ,  $q_1\bar{q}_1$  для AA, Aa, aa відповідно. Як і у випадку самозапліднення, стійкий розподіл досягається за одне покоління.

Проте, якщо розглянути статеві хромосоми, ситуація різко зміниться. Причина цього полягає в тому, що в деяких диплоїдних клітинах можуть знаходитися два різні типи хромосом X та Y. У ссавців жіночі особини характеризуються наявністю двох гомологічних X-хромосом, в той час як у чоловічих особин спарені хромосоми x і Y.

*Зауваження 1.* Гомологічними називаються хромосоми, що утворюють пару (в диплоїдному випадку). Всі нестатеві гомологічні хромосоми тождні.

Більше того, багато, якщо не всі, генів з X-хромосоми не мають complementів (від лат. complementum — доповнення) на Y-хромосомі. У випадку, що розглядається, генотип чоловічої особини повністю визначається алелями, локалізованими на X-хромосомі. Такі гени та ознаки, що ними контролюються, називаються *зчепленими зі статтю*.

Розглянемо розподіл зчепленого зі статтю гену в послідовних поколіннях. Припустимо, що початкові частоти генотипів у чоловічих та жіночих особин дорівнюють

Жіночі особини (XX)			Чоловічі особини (XY)	
AA	Aa	aa	A	a
$r_0$	$2s_0$	$t_0$	$p_0$	$q_0$

Помітимо, що генотип у чоловічої особини визначається лише одним символом гену, оскільки Y-хромосома не містить жодного гену, що мі-

ститься в  $X$ -хромосомі.

Частоти гамет у жіночих особин дорівнюють

$$\begin{array}{cc} A & a \\ r_0 + s_0 & s_0 + t_0 \end{array}$$

В припущенні про випадкове спарювання, або, що те саме, про випадковий вибір з сукупності гамет, частоти генотипів в наступному поколінні дорівнюють у жіночих особин

$$\begin{array}{ccc} AA & Aa & aa \\ r_1 = (r_0 + s_0)p_0 & 2s_1 = (r_0 + s_0)q_0 + (s_0 + t_0)p_0 & t_1 = (s_0 + t_0)q_0; \end{array}$$

та у чоловічих особин

$$\begin{array}{cc} A & a \\ p_1 = r_0 + s_0 & q_1 = s_0 + t_0. \end{array}$$

Кожна новостворена чоловіча зигота отримує  $Y$ -хромосому з чоловічої популяції та  $X$ -хромосому з жіночої. Отже частка генів  $A$  в чоловічій популяції в точності дорівнює частці генів  $A$  в попередньому поколінні жіночої популяції.

В загальному випадку якщо  $p_n$  — частота алеля  $A$  в сукупності жіночих гамет в  $n$ -му поколінні, то  $p_{n-1}$  — частота алеля  $A$  в чоловічій популяції в  $n$ -му поколінні. Отже,  $(n + 1)$ -ше покоління жіночих генотипів має склад

$$\begin{array}{ccc} AA & Aa & aa \\ p_n p_{n-1} & p_n q_{n-1} + q_n p_{n-1} & q_n q_{n-1}. \end{array}$$

Звідси можна вивести рекурентне співвідношення

$$p_{n+1} = p_n p_{n-1} + \frac{1}{2}(p_n q_{n-1} + q_n p_{n-1}),$$

яке зводиться до наступного

$$\begin{aligned} 2p_{n+1} &= 2p_n p_{n-1} + p_n(1 - p_{n-1}) + (1 - p_n)p_{n-1} \\ &= 2p_n p_{n-1} + p_n - p_n p_{n-1} + p_{n-1} - p_n p_{n-1} = p_n + p_{n-1}, \end{aligned}$$

тобто  $2[p_{n+1} - p_n] = -(p_n - p_{n-1})$ ,  
 або  $2\Delta_n = -\Delta_{n-1}$ , де  $\Delta_n = p_{n+1} - p_n$ .

За індукцією отримаємо

$$\Delta_n = \left(-\frac{1}{2}\right)^n \Delta_0,$$

або

$$p_{n+1} = p_n + \left(-\frac{1}{2}\right)^n \Delta_0.$$

Звідси

$$\begin{aligned} p_{n+1} &= \left(-\frac{1}{2}\right)^n \Delta_0 + \left(-\frac{1}{2}\right)^{n-1} \Delta_0 + \dots + \Delta_0 + p_0 \\ &= p_0 + \Delta_0 \left(\frac{1 - \left(-\frac{1}{2}\right)^{n+1}}{3/2}\right). \end{aligned}$$

В границі

$$\alpha = \lim p_n = p_0 + \frac{2}{3}(p_1 - p_0) = \frac{2}{3}p_1 + \frac{1}{3}p_0.$$

Отримали вираз для граничної частоти генів  $A$  в чоловічій популяції.  
 В жіночій популяції граничні частоти генотипів мають вигляд

$$\begin{array}{ccc} AA & Aa & aa \\ \alpha^2 & 2\alpha(1 - \alpha) & (1 - \alpha)^2. \end{array}$$

Це, звичайно, в точності відповідає частотам Харді-Вайнберга.

## 2. Ланцюги Маркова

### 2.1. Означення

Розглянемо випадковий процес  $X_t$ ,  $t \in T$ , простір станів якого злічений або скінчений, а множина індексів  $T = (0, 1, 2, \dots)$ . Будемо позначати його  $\{X_n\}$ . Часто простір станів процесу зручно ототожнювати з множиною невід'ємних цілих чисел  $(0, 1, 2, \dots)$  та казати, що  $X_n$  знаходиться в стані  $i$ , якщо  $X_n = i$ .

Випадковий процес  $\{X_n\}$  називається ланцюгом Маркова, якщо

$$P\{X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$$



Якщо  $P\{X_0 = i\} = p_i$ , то

$$P\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \dots P_{i_0, i_1} p_{i_0}.$$

Отже, процес повністю визначений, якщо задані однокрокові ймовірності переходу та розподіл ймовірностей станів процесу в момент часу 0 (тобто в.в.  $X_0$ ).

## 2.2. Матриці ймовірностей переходу ланцюга Маркова

Аналіз ланцюга Маркова пов'язаний головним чином з обчисленням ймовірностей можливих його реалізацій, найважливішою характеристикою яких є матриця ймовірностей переходів за  $n$  кроків  $P^{(n)} = (P_{ij}^n)$ .  $P_{ij}^n$  позначає ймовірність того, що процес перейде зі стану  $i$  в стан  $j$  за  $n$  переходів, тобто

$$P_{ij}^n = P\{X_{n+m} = j | X_m = i\}. \quad (2.1)$$

**Теорема 1.** *Якщо  $P = (P_{ij})$  — матриця однокрокових ймовірностей переходу ланцюга Маркова, то*

$$P_{ij}^n = \sum_{k=0}^{\infty} P_{ik}^r P_{kj}^s \quad (2.2)$$

для довільної фіксованої пари невід'ємних цілих чисел  $r$  та  $s$ , такої, що  $r + s = n$ . При цьому за означенням  $P_{ij}^0 = \begin{cases} 1, & \text{якщо } i = j \\ 0, & \text{якщо } i \neq j. \end{cases}$

В формулі (2.2) неважко впізнати формулу множення матриць. Звідси випливає, що  $P^{(n)} = P^n$ ; іншими словами, ймовірності  $P_{ij}^n$  можна розглядати як елементи матриці  $P^n$  —  $n$ -ої степені матриці  $P$ .

Якщо ймовірність того, що процес в початковий момент знаходиться в стані  $j$  дорівнює  $p_j$ , то ймовірність опинитися в стані  $k$  в момент  $n$  дорівнює

$$p_k^{(n)} = \sum_{j=0}^{\infty} p_j P_{jk}^n = P\{X_n = k\}.$$



### 2.3. Класифікація станів ланцюга Маркова

Стан  $j$  називається *досяжним* зі стану  $i$ , якщо  $P_{ij}^n > 0$  для деякого цілого числа  $n \geq 0$ , тобто ймовірність того, що процес за скінчену кількість кроків потрапить в стан  $j$ , вирушаючи зі стану  $i$ , додатня. Це позначають  $i \rightarrow j$ .

Стани  $i$  та  $j$  називаються *сполученими* ( $i \leftrightarrow j$ ), якщо  $i \rightarrow j$  та  $j \rightarrow i$ .

Властивість сполученості є відношенням еквівалентності, тобто 1)  $i \leftrightarrow j$ ; 2)  $i \leftrightarrow j \Rightarrow j \leftrightarrow i$ ; 3)  $i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k$ .

Тому всю множину станів можна розбити на класи еквівалентності. Стани об'єднуються в один клас, якщо вони сполучаються один з одним. Можливо, що вирушаючи зі стану, що належить одному класу, ми з додатньою ймовірністю потрапимо в інший клас, але тоді, очевидно, повернення в початковий клас вже неможливе, оскільки інакше обидва згаданих класи входили б в один клас еквівалентності.

Будемо казати, що ланцюг Маркова *нерозкладний*, якщо введене нами відношення еквівалентності породжує лише один клас станів. Іншими словами, процес нерозкладний, якщо всі його стани сполучаються один з одним.

### 2.4. Періодичність ланцюга Маркова

Визначимо період стану  $i$  як НСД всіх цілих чисел  $n \geq 1$ , для яких  $P_{ii}^n > 0$ . Якщо  $P_{ii}^n = 0$  при всіх  $n \geq 1$ , то  $d(i) := 0$ .

**Теорема 2.** *Якщо  $i \leftrightarrow j$ , то  $d(i) = d(j)$ .*

Тобто період є характеристикою класу сполучених станів.

**Теорема 3.** *Якщо стан  $i$  має період  $d(i)$ , то існує ціле число  $N$ , що залежить від  $i$ , таке, що для всіх цілих чисел  $n \geq N$ :  $P_{ii}^{nd(i)} > 0$ .*

Тобто повернення в стан  $i$  може відбуватися у всі досить далекі моменти часу, що кратні періоду  $d(i)$ .

**Наслідок 1.** Якщо  $P_{ij}^m > 0$ , то  $P_{ij}^{m+nd(i)} > 0$  для всіх досить великих додатних цілих чисел  $n$ .

Ланцюг Маркова, кожний стан якого має період 1, називається *неперіодичним*.

## 2.5. Рекурентність

Розглянемо довільний, але фіксований стан  $i$ . Покладемо для кожного цілого числа  $n \geq 1$ :

$$f_{ii}^n = P\{X_n = i, X_\nu \neq i, \nu = 1, 2, \dots, n-1 | X_0 = i\}.$$

Іншими словами,  $f_{ii}^n$  — це ймовірність того, що вирушаючи зі стану  $i$ , система вперше повернеться в цей стан через  $n$  переходів.

*Зауваження 2.* Через  $f_{ij}^k$  ( $i \neq j$ ) позначають ймовірність того, що процес вперше досягне стану  $j$  із стану  $i$  на  $k$ -му кроці.

Зрозуміло, що  $f_{ii}^1 = P_{ii}$ , а  $f_{ii}^n$  можна обчислити рекурентно з формули  $P_{ii}^n = \sum_{k=0}^{n-1} f_{ii}^k P_{ii}^{n-k}$ ,  $n \geq 1$ , де, за означенням,  $f_{ii}^0 = 0$  для всіх  $i$ .

*Зауваження 3.* Для  $i \neq j$ :  $P_{ij}^n = \sum_{k=0}^{n-1} f_{ij}^k P_{jj}^{n-k}$ ,  $n \geq 0$ , де  $f_{ij}^0 = 0$  для всіх  $i$  та  $j$ .

Будемо називати стан  $i$  *рекурентним*, якщо  $\sum_{n=1}^{\infty} f_{ii}^n = 1$ . Це означає, що стан  $i$  є рекурентним тоді і тільки тоді, коли ймовірність повернутися в початковий стан  $i$  після деякого скінченного числа кроків дорівнює 1.

Якщо  $\sum_{n=1}^{\infty} f_{ii}^n < 1$ , то стан  $i$  будемо називати *перехідним*.

**Теорема 4.** Стан  $i$  є рекурентним тоді і тільки тоді, коли  $\sum_{n=1}^{\infty} P_{ii}^n = \infty$ .

*Зауваження 4.* Якщо  $j$  — перехідний стан, то для всіх  $i$ :  $\sum_{n=1}^{\infty} P_{ij}^n < \infty$ .

**Наслідок 2.** Якщо  $i \leftrightarrow j$  та  $i$  — рекурентний стан, то стан  $j$  також є рекурентним.

Таким чином рекурентність, як і періодичність, є властивістю класу еквівалентності, тобто всі стани в класі еквівалентності або рекурентні, або перехідні одночасно.

Позначимо

$$Q_{ii} = P \left\{ \begin{array}{l} \text{частинка, вирушаючи зі стану } i, \\ \text{повертається в нього нескінченно часто} \end{array} \right\}.$$

**Теорема 5.** Стан  $i$  рекурентний або перехідний в залежності від того,  $Q_{ii} = 1$  або  $Q_{ii} = 0$ .

Тобто якщо деякий стан рекурентний, то цей стан з імовірністю 1 зустрічається в процесі нескінчену кількість разів.

**Теорема 6.** Якщо  $i \leftrightarrow j$  і обидва стани належать рекурентному класу, то

$$f_{ij}^* := \sum_{n=1}^{\infty} f_{ij}^n = 1.$$

Нехай

$$Q_{ij} = P \left\{ \begin{array}{l} \text{частинка, вирушаючи зі стану } i, \text{ буде} \\ \text{нескінченну кількість разів знаходитись в стані } j \end{array} \right\}.$$

**Наслідок 3.** Якщо  $i \leftrightarrow j$  і обидва стани належать рекурентному класу, то  $Q_{ij} = 1$ .

## 2.6. Основні граничні теореми для ланцюгів Маркова

Наступна теорема, так звана “ергодична” теорема, описує граничну поведінку ймовірностей  $P_{ij}^n$  при  $n \rightarrow \infty$  для всіх  $i$  та  $j$  неперіодичного рекурентного ланцюга Маркова.

**Теорема 7 (Основна гранична теорема для ланцюгів Маркова).** (а) Розглянемо рекурентний нерозкладний неперіодичний ланцюг Маркова. Нехай  $P_{ii}^n$  — ймовірність опинитися в стані  $i$  на  $n$ -му кроці,

$n = 0, 1, 2, \dots$ , при умові, що  $X_0 = i$  (тобто стан  $i$  – початковий). Нехай, як і раніше,  $P_{ii}^0 = 1$ . Нехай  $f_{ii}^n$  – ймовірність вперше повернутися в стан  $i$  на  $n$ -му кроці, причому  $f_{ii}^0 = 0$ . Як відомо,

$$P_{ii}^n - \sum_{k=0}^n f_{ii}^{n-k} P_{ii}^k = \begin{cases} 1, & \text{якщо } n = 0 \\ 0, & \text{якщо } n > 0. \end{cases}$$

В цьому випадку має місце рівність

$$\lim_{n \rightarrow \infty} P_{ii}^n = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^n}.$$

(б) При цих же умовах  $\lim_{n \rightarrow \infty} P_{ji}^n = \lim_{n \rightarrow \infty} P_{ii}^n$ .

*Зауваження 5.* Нехай  $C$  – рекурентний клас. Тоді  $P_{ij}^n = 0$  при  $i \in C, j \notin C$  для всіх  $n$ . Отже, потрапивши в  $C$ , вийти з нього неможливо. Таким чином, підматриця  $(P_{ij}), i, j \in C$  є матрицею ймовірностей переходу, а відповідний ланцюг маркова нерозкладний і рекурентний. Це означає, що гранична теорема має місце дослівно для будь-якого неперіодичного рекурентного класу.

*Зауваження 6.* Якщо  $a_n \rightarrow a$  при  $n \rightarrow \infty$ , то виконується рівність

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n a_k = a. \quad (2.3)$$

Отже, якщо стан  $i$  належить рекурентному неперіодичному класу, то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ii}^m = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^n} = \frac{1}{m_i}, \quad (2.4)$$

де  $m_i$  – середній час повернення.

Якщо стан  $i$  належить до рекурентного періодичного класу, то  $P_{ii}^m = 0$ , якщо  $m$  не кратне періоду  $d$  (тобто якщо  $m \neq nd$  для будь-якого  $n$ ), та

$$\lim_{n \rightarrow \infty} P_{ii}^{nd} = \frac{d}{m_i}.$$

Ці два останні результати разом з (2.3) показують, що співвідношен-

ня (2.4) має місце і для періодичного випадку.

Якщо  $\lim_{n \rightarrow \infty} P_{ii}^n = \pi_i > 0$  для деякого стану  $i$  з неперіодичного рекурентного класу, то  $\pi_j > 0$  для всіх  $j$  з цього класу. В цьому випадку називатимемо клас *позитивно рекурентним*, або *сильно ергодичним*. Якщо всі  $\pi_i = 0$  і клас рекурентний, то будемо казати, що клас *нуль-рекурентний*, або *слабко ергодичний*.

**Теорема 8.** Для неперіодичного позитивно рекурентного класу зі станами  $j = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P_{jj}^n = \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad \sum_{i=0}^{\infty} \pi_i = 1,$$

та величини  $\{\pi_i\}$  однозначно визначаються умовами

$$\pi_i \geq 0, \quad \sum_{i=0}^{\infty} \pi_i = 1, \quad \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}. \quad (2.5)$$

Набір  $\{\pi_i\}$ , що задовольняє умовам (2.5), називається *стаціонарним розподілом* ланцюга Маркова.

*Зауваження 7.* Якщо  $P\{X_0 = j\} = \pi_j$ , то  $P\{X_n = j\} = \pi_j$  для всіх  $n$ .

## 2.7. Імовірності поглинання

Як відомо з попереднього, якщо стан  $j$  перехідний, то  $P_{ij}^n \rightarrow 0$ , і якщо стани  $i$  та  $j$  належать до одного й того ж неперіодичного рекурентного класу, то  $P_{ij}^n \rightarrow \pi_j \geq 0$ . Якщо стани  $i$  та  $j$  входять в один і той же рекурентний періодичний клас, то останнє твердження зберігає силу, якщо  $P_{ij}^n$  замінити в ньому на  $n^{-1} \sum_{m=1}^n P_{ij}^m$ . Для того, щоб завершити розгляд граничної поведінки ймовірностей  $P_{ij}^n$ , залишається розглянути випадок, коли стан  $i$  перехідний, а стан  $j$  рекурентний.

Нехай  $T$  — множина всіх перехідних станів; введемо величини  $x_i^n$  за допомогою рекурентної формули

$$x_i^1 = \sum_{j \in T} P_{ij},$$

$$x_i^n = \sum_{j \in T} P_{ij} x_j^{n-1},$$

де  $i \in T$  та  $n \geq 2$ . Помітимо, що  $x_i^n$  — це ймовірність того, що, вирушивши зі стану  $i$ , процес не вийде з класу  $T$  протягом наступних  $n$  кроків. Покажемо за допомогою індукції, що послідовність  $\{x_i^n, n = 1, 2, \dots\}$  є незростаючою.

Дійсно, оскільки  $x_i^n \leq 1$  при всіх  $n$ , то

$$x_i^2 = \sum_{j \in T} P_{ij} x_j^1 \leq \sum_{j \in T} P_{ij} = x_i^1.$$

Припустимо тепер, що  $x_j^n \leq x_j^{n-1}$  при всіх  $j \in T$ , тоді

$$0 \leq x_i^{n+1} = \sum_{j \in T} P_{ij} x_j^n \leq \sum_{j \in T} P_{ij} x_j^{n-1} = x_i^n.$$

Це означає, що обмежена послідовність  $\{x_i^n, n = 1, 2, \dots\}$  не зростає і, отже, збігається до деякої границі  $x_i$ , причому

$$x_i = \sum_{j \in T} P_{ij} x_j, \quad i \in T. \quad (2.6)$$

Таким чином, якщо єдиним обмеженням розв'язком рівняння (2.6) є нульовий вектор  $(0, 0, \dots)$ , то вирушаючи з будь-якого перехідного стану, процес з імовірністю 1 буде поглинутий деяким класом рекурентних станів. Справді,  $x_i$  ( $i \in T$ ) — це ймовірність ніколи не потрапити в рекурентний клас, якщо  $i$  був початковим станом процесу. Оскільки  $\{x_i, i \in T\}$  є обмеженим розв'язком рівнянь (2.6), то  $x_i = 0$  при всіх  $i$ .

*Зауваження 8.* Якщо ланцюг Маркова має лише скінчену кількість станів, наприклад  $M$ , то серед них немає нуль-рекурентних станів, а всі стани не можуть бути перехідними. Дійсно, оскільки  $\sum_{j=0}^{M-1} P_{ij}^n = 1$  для всіх  $n$ , то  $\lim_{n \rightarrow \infty} P_{ij}^n = 0$  не для всіх  $j$ . З цієї ж причини відсутні нуль-рекурентні стани.

Нехай  $C, C_1, C_2, \dots$  — рекурентні класи. Визначимо  $\pi_i(C)$  як імовірність того, що вирушаючи з перехідного стану  $i$ , процес рано чи пізно увійде в клас  $C$ . (Згадаємо, що одного разу потрапивши в рекурентний

клас, процес вже ніколи його не залишає).

Нехай  $\pi_i^n(C)$  — ймовірність того, що процес досягне класу  $C$  і, отже, буде ним поглинутий, вперше на  $n$ -му кроці, при умові, що початковим станом було  $i \in T$ ; тоді

$$\pi_i(C) = \sum_{n=1}^{\infty} \pi_i^n(C) \leq 1, \quad (2.7)$$

$$\pi_i^1(C) = \sum_{j \in C} P_{ij},$$

$$\pi_i^n(C) = \sum_{j \in T} P_{ij} \pi_j^{n-1}(C), \quad n \geq 2. \quad (2.8)$$

Використовуючи (2.8), (2.7) можна переписати у вигляді

$$\begin{aligned} \pi_i(C) &= \pi_i^1(C) + \sum_{n=2}^{\infty} \pi_i^n(C) = \pi_i^1(C) + \sum_{n=2}^{\infty} \sum_{j \in T} P_{ij} \pi_j^{n-1}(C) \\ &= \pi_i^1(C) + \sum_{j \in T} P_{ij} \sum_{n=2}^{\infty} \pi_j^{n-1}(C), \\ \pi_i(C) &= \pi_i^1(C) + \sum_{j \in T} P_{ij} \pi_j(C), \quad i \in T. \end{aligned} \quad (2.9)$$

Якщо припустити, що єдиним обмеженням розв'язком однорідної системи рівнянь

$$w_i = \sum_{j \in T} P_{ij} w_j, \quad i \in T,$$

є тривіальний розв'язок (нульовий вектор), то  $\{\pi_i(C)\}$  є єдиним обмеженням розв'язком системи рівнянь (2.9). Більше того, або  $\pi_i^1(C) > 0$  для деякого  $i \in T$ , або ж  $\pi_i(C) = 0$  для всіх  $i \in T$  і, отже,  $\pi_i^n(C) = 0$  для всіх  $n$ .

**Теорема 9.** *Нехай  $j \in C$  ( $C$  — неперіодичний рекурентний клас), тоді для  $i \in T$  маємо*

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_i(C) \lim_{n \rightarrow \infty} P_{jj}^n = \pi_i(C) \pi_j.$$

*Зауваження 9.* Якщо  $C$  — періодичний клас, та  $j \in C$ , то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^m = \pi_i(C)\pi_j.$$

Наостанок помітимо, що якщо  $i$  — перехідний стан, а  $j$  — рекурентний, то граничне значення ймовірності залежить від обох станів  $i$  та  $j$ . В цьому полягає суттєва відмінність від випадку, коли  $i$  та  $j$  належать одному й тому ж рекурентному класу.

### 3. Інбридинг

*Інбридингом* в біології називається спарювання двох близькоспоріднених індивідуумів.

Існує два типи невідповідного спарювання. Перший з них, що називається *асортативним спарюванням* передбачає обов'язковість спарювання подібних (або неподібних) генотипів (або фенотипів). Очевидним прикладом є спарювання генотипів  $XX$  та  $XY$  при статевому розмноженні. Іншим типом невідповідного спарювання є спарювання близькоспоріднених індивідуумів.

#### 3.1. Самозапліднення

Перша модель інбридингу, яку ми розглянемо, відноситься до самозапліднення. Кожен індивідуум при цьому має обидві гамети. Допускається лише спарювання типів  $AA \times AA$ ,  $aa \times aa$ ,  $Aa \times Aa$ , тобто спарюватись можуть лише індивідууми, які мають один і той же генотип.

Наприклад, пиляк у квітів виділяє пилок (чоловічі гамети), а в маточці розвиваються сім'яруньки (жіночі гамети) і при цьому часто відбувається свого роду інбридинг.

Якщо початкові частоти зигот  $AA$ ,  $Aa$ ,  $aa$  дорівнюють відповідно  $u_0$ ,  $v_0$ ,  $w_0$ , то аналогічні частоти для наступного покоління дорівнюють

$$u_1 = u_0 + \frac{1}{4}v_0; \quad v_1 = \frac{1}{2}v_0; \quad w_1 = w_0 + \frac{1}{4}v_0,$$



оскільки при спарюванні  $Aa \times Aa$  утворюються нащадки в співвідношенні  $\frac{1}{4}AA, \frac{1}{2}Aa, \frac{1}{4}aa$ . З кожним наступним поколінням частота гетерозигот  $Aa$  зменшується в 2 рази. В  $n$ -му поколінні частоти дорівнюють

$$u_n = u_0 + \frac{1}{4}v_0 \left[ 1 + \frac{1}{2} + \dots + \left(\frac{1}{2}\right)^{n-1} \right] = u_0 + \frac{1}{2}v_0 \left[ 1 - \left(\frac{1}{2}\right)^n \right],$$

$$v_n = \left(\frac{1}{2}\right)^n v_0,$$

$$w_n = w_0 + \frac{1}{4}v_0 \left[ 1 + \frac{1}{2} + \dots + \left(\frac{1}{2}\right)^{n-1} \right] = w_0 + \frac{1}{2}v_0 \left[ 1 - \left(\frac{1}{2}\right)^n \right].$$

Граничні частоти при  $n \rightarrow \infty$  дорівнюють:

$$u = u_0 + \frac{v_0}{2}, \quad v = 0, \quad w = w_0 + \frac{v_0}{2}.$$

### 3.2. Спарювання сібсів

В наступній моделі розглядається спарювання нащадків одних і тих же батьків, або спарювання сібсів (sibs = sisters/brothers). Знову метою є визначення граничних частот різних генотипів, що приймають участь у спарюванні. Однак замість безпосереднього розгляду частот генотипів ми прослідкуємо за флуктуаціями частот особин, які спарюються, в послідовних поколіннях.

Якщо обмежитись розглядом двох алелей  $A$  і  $a$ , то існують три можливі генотипи для кожного з батьків та шість можливих варіантів поєднання, або схрещування генотипів, тобто існує 6 різних варіантів спарювання:  $AA \times AA, aa \times aa, AA \times Aa, aa \times Aa, Aa \times Aa, AA \times aa$ . Припустимо, що частоти кожного з цих варіантів відомі. Ми хочемо обчислити ці частоти для наступного покоління. Для того, щоб зробити це, ми зараз знайдемо розподіл генотипів нащадків, а потім, припускаючи, що спарюватись можуть лише сібси, визначимо частоту різних видів спарювання нащадків, що отримуються при цьому.

Наприклад, всі діти, що отримуються при спарюванні  $AA \times AA$ , мають тип  $AA$ , і тому серед сібсів допустимим є лише спарювання вигляду  $AA \times AA$ . При спарюванні  $AA \times Aa$  утворюються нащадки у відношенні

$\frac{1}{2}AA, \frac{1}{2}Aa$ . Частоти різних варіантів спарювання серед сібсів, очевидно, дорівнюють  $\frac{1}{4}AA \times AA, \frac{1}{2}AA \times Aa, \frac{1}{4}Aa \times Aa$ . Аналогічним чином можна розглянути й інші можливості. Так, при спарюванні  $Aa \times Aa$  нащадки утворюються у відношенні  $\frac{1}{4}AA, \frac{1}{4}Aa, \frac{1}{4}aa$ . Частоти різних варіантів спарювання серед сібсів можна легко обчислити. Вони дорівнюють  $\frac{1}{16}AA \times AA, \frac{1}{16}aa \times aa, \frac{1}{4}Aa \times Aa, \frac{1}{8}AA \times aa, \frac{1}{4}AA \times Aa, \frac{1}{4}aa \times Aa$ .

Інші поєднання досліджуються аналогічним чином.

Можна проаналізувати зміни частот типів, що спарюються, від покоління до покоління за допомогою ланцюгів Маркова. Для цього розглянемо ланцюг Маркова з 6 станами, які відповідають 6 варіантам спарювання: стан 1 —  $AA \times AA$ , стан 2 —  $AA \times Aa$ , стан 3 —  $AA \times aa$ , стан 4 —  $Aa \times Aa$ , стан 5 —  $Aa \times aa$ , стан 6 —  $aa \times aa$ .

Цей ланцюг Маркова має наступну матрицю переходу:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

В рядочках матриці записані частоти різних варіантів спарювання серед сібсів, які народилися в результаті відповідного номеру рядочку варіанту спарювання. Є два поглинаючі стани: 1 і 6.

З теорії ланцюгів Маркова відомо, що в решті решт відбудеться поглинання в одному з цих двох станів. Імовірність поглинання з перехідних станів (стани 2 — 5) цікавить нас, оскільки вона, очевидно, відповідає тій інтенсивності, з якою зникають гетерозиготні генотипи. Ця інтенсивність дорівнює, звичайно, найбільшому власному значенню ма-

триці

$$B = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

яка одержиться при виключенні першого та шостого станів. Оскільки наслідки варіантів спарювання  $AA \times Aa$  та  $aa \times Aa$  симетричні, можна належним чином об'єднати перший і четвертий рядочки матриці  $B$ . При цьому отримаємо матрицю

$$C = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{4} \end{pmatrix}.$$

Власні значення  $C$  є коренями рівняння

$$\begin{vmatrix} \frac{1}{2} - \lambda & 0 & \frac{1}{4} \\ 0 & -\lambda & 1 \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{4} - \lambda \end{vmatrix} = 0, \text{ або}$$

$$\left(\frac{1}{2} - \lambda\right) \left(\lambda^2 - \frac{1}{4}\lambda - \frac{1}{8}\right) + \frac{1}{8}\lambda = 0;$$

$$(1 - 2\lambda)(8\lambda^2 - 2\lambda - 1) + 2\lambda = 0;$$

$$-16\lambda^3 + 12\lambda^2 + 2\lambda - 1 = 0;$$

$$\begin{array}{c|c|c|c|c} & -16 & 12 & 2 & -1 \\ \hline \frac{1}{4} & -16 & 8 & 4 & 0 \end{array}$$

$$(4\lambda - 1)(4\lambda^2 - 2\lambda - 1) = 0;$$

$$\lambda_1 = \frac{1}{4}, \quad \lambda_2 = \frac{1 - \sqrt{5}}{4}, \quad \lambda_3 = \frac{1 + \sqrt{5}}{4}.$$

Отже швидкість наближення до гомозиготного стану (фіксації) до-

рівнює  $\lambda_3 = \frac{1 + \sqrt{5}}{4}$ , тобто з кожним поколінням частота варіантів спарювання різних генотипів зменшується в  $\lambda_3$  разів.

Не складно обчислити при будь-яких початкових частотах  $(u_1, u_2, \dots, u_6)$  варіантів спарювання фінальну частоту поглинання, або фіксації, в станах 1 ( $AA \times AA$ ) або 6 ( $aa \times aa$ ).

Ймовірність  $x_i$  фіксації в стані 1 (спарювання  $AA \times AA$ ) при початковому стані  $i$ ,  $i = 1, 2, \dots, 6$ , є розв'язком рівняння

$$\sum_{j=1}^6 a_{ij} x_j = x_i,$$

де покладається  $x_1 = 1$ ,  $x_6 = 0$  і  $A = (a_{ij})$ .

$$\begin{cases} \frac{1}{4} + \frac{1}{2}x_2 + \frac{1}{4}x_4 = x_2, \\ x_4 = x_2, \\ \frac{1}{16} + \frac{1}{4}x_2 + \frac{1}{8}x_3 + \frac{1}{4}x_4 + \frac{1}{4}x_5 = x_4, \\ \frac{1}{4}x_4 + \frac{1}{2}x_5 = x_5; \\ x_2 = \frac{3}{4}, \quad x_3 = x_4 = \frac{1}{2}, \quad x_5 = \frac{1}{4}. \end{cases}$$

Отже,  $x_1 = 1$ ,  $x_2 = \frac{3}{4}$ ,  $x_3 = \frac{1}{2}$ ,  $x_4 = \frac{1}{2}$ ,  $x_5 = \frac{1}{4}$ ,  $x_6 = 0$ .

Якщо початкові частоти різних станів дорівнюють  $(u_1, u_2, \dots, u_6)$ , то ймовірність фіксації в стані 1, очевидно, дорівнює

$$\sum_{i=1}^6 u_i x_i = u_1 + \frac{3}{4}u_2 + \frac{1}{2}u_3 + \frac{1}{2}u_4 + \frac{1}{4}u_5.$$

### 3.3. Гени, ідентичні за походженням

Ще одне визначення ступені інбридингу в популяції дається через поняття коефіцієнту інбридингу. *Коефіцієнт інбридингу* індивідуума  $X$  визначається як ймовірність того, що два гени в одному локусі у індивідуума  $X$  *ідентичні за походженням*. Під цим ми розуміємо, що два гени є “копіями” одного й того ж гену деякого індивідууму з попереднього покоління.

Розглянемо тепер скінчену популяцію, що складається з  $N$  індивідуумів, які відтворюються нестатевим шляхом, тобто кожен індивідуум “постачає” нескінчене число копій своїх двох генів в сукупність гамет, в якій потім відбувається випадкове спарювання з ціллю утворення нових  $N$  індивідуумів. Це означає, що кожен індивідуум постачає одне і те ж число генів  $k$  і розглядається граничний випадок  $k \rightarrow \infty$ . При цьому поняття частоти не втрачає змісту.

Нехай  $F_n$  — коефіцієнт інбридингу в  $n$ -му поколінні популяції, тобто  $F_n$  — ймовірність того, що два гени, вибрані випадковим чином, ідентичні за походженням.

В сукупності гамет відбувається їх спарювання з метою утворення  $N$  нових індивідуумів. Припустимо, що вибрана деяка гамета і треба випадковим чином вибрати другу гамету з тим, щоб спарувати з першою. Ймовірність того, що ген, який міститься в другій гаметі, буде копією того ж самого гена, що й у першій, дорівнює  $\frac{1}{2N}$ , оскільки кожен з  $2N$  типів гамет, вироблених  $N$  індивідуумами, є однаково ймовірним. При цьому два спарених гени є, звичайно, ідентичними за походженням.

Тепер, якщо два спарених гени не є копіями одного й того ж гену (ймовірність чого дорівнює  $1 - \frac{1}{2N}$ ), ймовірність того, що вони ідентичні за походженням, дорівнює  $F_n$ . Отже,

$$\begin{aligned}
 F_{n+1} &= \frac{1}{2N} + \frac{2N-1}{2N} F_n, \\
 1 - F_{n+1} &= 1 - \frac{1}{2N} - \frac{2N-1}{2N} F_n = \left(1 - \frac{1}{2N}\right) (1 - F_n), \\
 H_n &= 1 - F_n = \left(1 - \frac{1}{2N}\right) H_{n-1} \\
 &= \left(1 - \frac{1}{2N}\right)^2 H_{n-2} = \dots = \left(1 - \frac{1}{2N}\right)^n H_0.
 \end{aligned}$$

Таким чином, ймовірність того, що деякий індивідуум в  $n$ -му поколінні — гетерозигот, прямує до 0 при  $n \rightarrow \infty$  зі швидкістю  $\left(1 - \frac{1}{2N}\right)^n$ . Оскільки було зроблено припущення про великий розмір популяції, ця

швидкість досить низька. Значення проведеного аналізу полягає в отриманні якісного результату про те, що скінченність популяції тягне за собою (в деякій мірі) інбридинг.

### 3.4. Нескінченна диплоїдна популяція та коефіцієнт інбридингу

Припустимо, що є нескінченна диплоїдна популяція, у якій в одному локусі ген  $A$  може знаходитися з частотою  $p$ , а ген  $a$  — з частотою  $q = 1 - p$ . Припустимо, що популяція має сталий коефіцієнт інбридингу  $F$ , який має той самий зміст, що й раніше. Частка  $F$  популяції складається з гомозигот, у яких гени ідентичні за походженням. Очевидно, частки гомозигот  $AA$  та  $aa$  з генами, ідентичними за походженням, дорівнюють відповідно  $Fp$  та  $Fq$ . Частка всіх зигот, в яких гени не ідентичні за походженням, дорівнюють  $1 - F$ . Отже, частки типів  $AA$ ,  $Aa$  та  $aa$  відповідно дорівнюють  $(1 - F)p^2$ ,  $(1 - F)2pq$  та  $(1 - F)q^2$ . Тому, якщо  $u$ ,  $v$  і  $w$  — частоти генотипів  $AA$ ,  $Aa$  та  $aa$  відповідно, то

$$\begin{aligned} u &= (1 - F)p^2 + Fp = p^2 + Fpq, \\ v &= (1 - F)2pq, \\ w &= (1 - F)q^2 + Fq = q^2 + Fpq. \end{aligned} \tag{3.1}$$

Очевидно,  $u + v + w = 1$ , як і очікувалося. Ці частоти є модифікаціями частот Харді-Вайнберга. Помітимо, що частота гену  $A$  дорівнює  $u + \frac{1}{2}v = p^2 + pq = p$  і залишається сталою, якщо  $F = \text{const}$ . Значення коефіцієнта  $F$  стає зрозумілим, якщо помітити, що при  $F = 0$   $u$ ,  $v$ ,  $w$  є частотами Харді-Вайнберга, які описують популяцію без інбридингу, в той час як при  $F = 1$  частота гетерозиготних індивідуумів дорівнює нулю, тобто популяція досягає стану максимального інбридингу.

### 3.5. Популяція з самозаплідненням та випадковим спарюванням

Припустимо, що частка  $s$  популяції відтворюється самозаплідненням, в той час коли в іншій її частині має місце випадкове спарювання. Складемо таблицю з відповідними схемами розмноження.

	$AA$	$Aa$	$aa$
Частоти	$u$	$v$	$w$
Нащадки від самозапліднення	$su(AA)$	$sv \left( \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa \right)$	$sw(aa)$
Нащадки від випадкового спарювання	$(1-s)p^2(AA)$	$(1-s)2pq(Aa)$	$(1-s)q^2(aa)$

Тут  $p = u + \frac{1}{2}v$ ,  $q = 1 - p = w + \frac{1}{2}v$ .

Частоти генотипів в другому поколінні дорівнюють

$$u' = (1-2)p^2 + su + \frac{1}{4}sv,$$

$$v' = 2(1-s)pq + \frac{1}{2}sv,$$

$$w' = (1-s)q^2 + sw + \frac{1}{4}sv.$$

Крім того,

$$p' = u' + \frac{1}{2}v' = (1-s)p + s \left( u + \frac{1}{2}v \right) = (1-s)p + sp = p.$$

Таким чином,  $p$  залишається сталим від покоління до покоління, як і слід було очікувати, оскільки ми не розглядаємо ніяких факторів, які б могли впливати на частоту гену.

Отже, можна записати

$$\begin{aligned} v_{n+1} &= 2(1-s)pq + \frac{1}{2}sv_n \\ &= 2(1-s)pq \left( 1 + \frac{1}{2}s + \left( \frac{1}{2}s \right)^2 + \dots + \left( \frac{1}{2}s \right)^n \right) + \left( \frac{1}{2}s \right)^{n+1} v_0 \\ &= \frac{4(1-s)}{2-s}pq \left[ 1 - \left( \frac{1}{2}s \right)^{n+1} \right] + \left( \frac{1}{2}s \right)^{n+1} v_0, \end{aligned}$$

$$\begin{aligned}
u_n &= (1-s)p^2 + s \left( u_{n-1} + \frac{1}{4}v_{n-1} \right) = \left| \text{оскільки } u_{n-1} + \frac{1}{2}v_{n-1} = p \right| \\
&= (1-s)p^2 + s \left( p - \frac{1}{4}v_{n-1} \right) = p^2 + pqs - \frac{1}{4}sv_{n-1} \\
&= p^2 + pqs - \frac{1}{4}s \frac{4(1-s)}{2-s} pq \left[ 1 - \left( \frac{1}{2}s \right)^{n-1} \right] - \frac{1}{4}s \left( \frac{1}{2}s \right)^{n-1} v_0 \\
&= p^2 + pqs \left[ 1 - \frac{1-s}{2-s} \left( 1 - \left( \frac{1}{2}s \right)^{n-1} \right) \right] - \frac{1}{4}s \left( \frac{1}{2}s \right)^{n-1} v_0.
\end{aligned}$$

Отже, граничне значення  $u$  при  $n \rightarrow \infty$  дорівнює

$$p^2 + pqs \frac{1}{2-s} = p^2 + pq \frac{s}{2-s}.$$

Аналогічно, граничні значення  $v$  і  $w$  дорівнюють відповідно  $2pq \left( 1 - \frac{s}{2-s} \right)$  і  $q^2 + pq \frac{s}{2-s}$ . Таким чином, звертаючись до (3.1), ми бачимо, що в рівноважному стані  $F$  має значення  $\frac{s}{2-s}$ . Чим більше  $s$ , тим повільніше досягається рівноважний стан, і навпаки.

Зокрема, при  $s = 0$ , тобто при повністю випадковому спарюванні, коефіцієнт  $F$  дорівнює нулю і інбридингу немає. Якщо  $s = 1$ , або відбувається лише самозапліднення, то  $F = 1$  і в рівноважному стані популяція повністю гомозиготна.

#### 4. Поліплоїди

Багато організмів, особливо рослини, є поліплоїдами, тобто гомологічні хромосоми зустрічаються групами розміру більше ніж 2. Розглянемо  $2n$ -плоїдну популяцію (тобто набори гомологічних хромосом мають розмір  $2n$ ), у якої в кожному локусі, що розглядається, є лише два алелі. Будемо позначати комбінації зигот (генотипи) через  $A^s a^{2n-s}$  ( $s = 0, 1, \dots, 2n$ ), де  $s$  — число хромосом, що несуть ген  $A$ , а інші  $2n - s$  хромосом несуть ген  $a$ . Можливі гаметичні комбінації мають вигляд

$$A^r a^{n-r}, \quad r = 0, 1, \dots, n.$$

Розщеплення гамет в поліплоїді більш складне, ніж у випадку дипло-



їдних організмів. Це пояснюється тим фактом, що під час мейотичного циклу відбувається процес подвоєння, коли кожна хромосома розщеплюється вздовж своєї довжини на дві однакові хроматиди, які, проте, залишаються з'єднаними в пари за допомогою центрального стрижня — центромери. Між хроматидами гомологічних хромосом має місце деякий обмін генетичним матеріалом. Це називають явищем *кросинговера*. Природа розщеплення гамет залежить від відстані локусу, що розглядається, від центромери. Два найбільш розповсюджених способи розщеплення — хромосомне та хроматидне.

У випадку хромосомного розщеплення кожен з батьків  $A^s a^{2n-s}$  буде виробляти гамети  $A^r a^{n-r}$  з частотою

$$\frac{1}{C_{2n}^n} C_s^r C_{2n-s}^{n-r}, \quad r = 0, 1, \dots, s,$$

а у випадку хроматидного розщеплення з частотою

$$\frac{1}{C_{4n}^{2n}} C_{2s}^r C_{4n-2s}^{n-r}, \quad r = 0, 1, \dots, s.$$

У випадку  $n = 1$  ці два вирази співпадають.

Отримаємо аналог закону Харді-Вайнберга для локусу при хромосомному розщепленні. Якщо в популяції з  $n + 1$  гамет, в якій частота генів  $A$  дорівнює  $p$ , а частота генів  $a$  дорівнює  $q$ , відбувається випадкове злиття гамет, то частота особин  $A^s a^{2n-s}$  в наступному поколінні дорівнює  $C_{2n}^{2s} p^s q^{2n-s}$ . Відповідні гамети  $A^r a^{n-r}$  мають частоту

$$\begin{aligned} & \sum_{s=r}^{n+r} C_s^r C_{2n-s}^{n-r} (C_{2n}^n)^{-1} C_{2n}^s p^s q^{2n-s} \\ &= \sum_{s=r}^{n+r} \frac{s!}{r!(s-r)!} \frac{(2n-s)!}{(n-r)!(n+r-s)!} \frac{(n!)^2}{2n!} \frac{2n!}{s!(2n-s)!} p^s q^{2n-s} \\ &= \sum_{s=r}^{n+r} \frac{n!}{r!(n-r)!} p^r q^{n-r} \frac{n!}{(s-r)!(n-s+r)!} p^{s-r} q^{n-s+r} \\ &= C_n^r p^r q^{n-r} \sum_{s=r}^{n+r} C_n^{s-r} p^{s-r} q^{n-s+r} = C_n^r p^r q^{n-r} (p+q)^n = C_n^r p^r q^{n-r}. \end{aligned}$$

Таким чином, ген  $A$ , наприклад, має частоту

$$\sum_{r=0}^n \frac{r}{n} C_n^r p^r q^{n-r} = p \sum_{r=1}^n C_{n-1}^{r-1} p^{r-1} q^{n-r} = p.$$

Можна сформулювати наступне твердження про збіжність: нехай  $p_k(A^r a^{n-r})$ ,  $r = 0, 1, \dots, n$ , — частота гамет  $A^r a^{n-r}$  в  $k$ -му поколінні. Розглянемо випадкове спарювання і хромосомне розщеплення. Тоді

$$\lim_{k \rightarrow \infty} p_k(A^r a^{n-r}) = C_n^r p^r (1-p)^{n-r},$$

де  $p$  — початкова частота алеля  $A$  в популяції. Збіжність має швидкість геометричної прогресії із знаменником  $\frac{n-1}{2n-1}$ , тобто

$$|p_k(A^r a^{n-r}) - C_n^r p^r (1-p)^{n-r}| \leq C \lambda^k, \quad r = 1, 2, \dots, n,$$

для всіх  $k$ , де  $\lambda = \frac{n-1}{2n-1}$ .

Ми не будемо доводити цю теорему. Її можна пояснити за допомогою наступних евристичних міркувань.

Нехай  $B_r = C_n^r p^r (1-p)^{n-r}$ . Для утворення гаметі оберемо  $n$  хромосом із сукупності розміру  $2n$ . Якщо  $n$  хромосом взяти від різних предків, то має місце хромосомне розщеплення нульового покоління популяції і частоти, що розглядаються, є біноміальними, тобто  $p_k(A^r a^{n-r}) = B_r$ . Ми, звичайно, очікуємо, що при великих  $k$  це відповідає дійсності.

Але припустимо тепер, що  $n-2$  хромосоми отримані від різних предків, а дві — від одного й того ж предка (інші можливості можна не враховувати, оскільки вони менш імовірні).

Нехай  $1 - \rho^{(k)}$  — ймовірність того, що всі  $n$  генів в гаметі  $k$ -го покоління отримані від різних предків нульового покоління. Тоді

$$p_k(A^r a^{n-r}) = (1 - \rho^{(k)}) B_r + \rho^{(k)} C_r,$$

де  $C_r$  не залежить від  $k$ , а залежить лише від кількості “споріднених” хромосом в гаметі. Отже,

$$p_k(A^r a^{n-r}) = B_r + \rho^{(k)}(C_r - B_r).$$

Для того, щоб отримати рекурентне співвідношення відносно  $\rho^{(k+1)}$ ,

помітимо, що  $\rho^{(k+1)}$  залежить від імовірності спарювання між гаметою, яка має хромосоми, що прийшли від різних предків, і гаметою, в якій є хоча б дві “споріднені” хромосоми. Але ми можемо розпорядитися довільним чином лише  $(n - 2)$  хромосомами — дві “споріднені” повинні залишитися в парі. Ми нехтуємо випадком, коли наявна більша кількість “споріднених” хромосом. Тоді

$$\rho^{(k+1)} = 2\rho^{(k)} \left(1 - \rho^{(k)}\right) \frac{C_2^2 C_{2n-2}^{n-2}}{C_{2n}^n}.$$

Нехтуючи в правій частині членом, що містить  $\left[\rho^{(k)}\right]^2$ , отримаємо

$$\rho^{(k+1)} = 2\rho^{(k)} \frac{C_2^2 C_{2n-2}^{n-2}}{C_{2n}^n} = 2\rho^{(k)} \frac{1}{2} \frac{n-1}{2n-1} = \rho^{(k)} \frac{n-1}{2n-1}.$$

Таким чином,  $p_k(A^r a^{n-r}) = B_r + \rho^{(k)}(C_r - B_r)$  збігається до  $B_r$ , оскільки  $\rho^{(k)} \rightarrow 0$  геометрично з параметром  $\frac{n-1}{2n-1}$ .

*Зуваження 10.* Оскільки

$$\rho^{(k+1)} \leq 2\rho^{(k)} \frac{C_2^2 C_{2n-2}^{n-2}}{C_{2n}^n},$$

то справжні значення  $\rho^{(k)}$  не перевищують розв’язку “лінеаризованої” системи. Зокрема, якщо розв’язок лінеаризованої системи збігається до 0 при  $k \rightarrow \infty$ , то це ж має місце і для вихідної системи.

## 5. Гіллясті процеси

Гіллясті процеси є прикладом ланцюгів Маркова. Вони виникають природньо в різних наукових дисциплінах. Найбільш цікавими для нас є наступні приклади.

**(а) Вживання прізвищ.** Прізвище успадковують лише сини. Припустимо, що кожен індивідуум з імовірністю  $p_k$  має  $k$  нащадків чоловічої статі. Далі, кожен індивідуум породжує 1-е, 2-е,  $\dots$ ,  $n$ -те покоління нащадків. Можна досліджувати розподіл такої випадкової величини, як число нащадків в  $n$ -му поколінні, або ймовірність того, що прізвище зникне. Такі питання вивчатимуться в даному розділі при загальному

аналізі гіллястих процесів.

**(б) Виживання мутантних генів.** Кожен окремий ген має можливість породити  $k$  “нащадків”,  $k = 1, 2, \dots$ , які є генами того ж типу. Проте будь-який ген також може трансформуватись в інший тип, що називається мутантним геном, який може стати першим в послідовності поколінь мутантних генів. Цікавить імовірність виживання мутантних генів в популяції вихідних генів.

Наведені приклади мають спільну структуру, що описується наступною моделлю.

### 5.1. Означення

Припустимо, що організм в кінці свого життя виробляє випадкове число  $\xi$  нащадків згідно розподілу ймовірностей

$$P\{\xi = k\} = p_k, \quad k = 0, 1, 2, \dots, \quad (5.1)$$

де  $p_k \geq 0$ ,  $k = 0, 1, 2, \dots$  та  $\sum_{k=0}^{\infty} p_k = 1$ . В свою чергу нащадки незалежно один від одного в кінці свого життя (для простоти тривалість життя вважається однаковою для всіх організмів) виробляють нащадків, кожен у відповідності з розподілом (5.1), продовжуючи таким чином свій біологічний вид. Процес  $\{X_n\}_{n \geq 0}$ , де  $X_n$  — чисельність популяції в  $n$ -му поколінні, є ланцюгом Маркова.

Справді, в скінченновимірному розподілі випадкових величин  $X_{n_1}, X_{n_2}, \dots, X_{n_r}, X_n$ ,  $n_1 < n_2 < \dots < n_r < n$ , істотними є лише дані останнього перепису популяції, оскільки число нащадків є функцією лише числа безпосередніх предків. Матриця перехідних імовірностей, очевидно, визначається співвідношенням

$$P_{ij} = P\{X_{n+1} = j | X_n = i\} = P\{\xi_1 + \dots + \xi_i = j\}, \quad (5.2)$$

де  $\xi_1, \dots, \xi_i$  — незалежні спостереження в.в., яка має розподіл (5.1). Формулу (5.2) можна пояснити наступним чином. Незалежно один від одного  $i$  індивідумів, які утворюють  $n$ -те покоління, виробляють на-

щадків ( $l$ -ий — в кількості  $\xi_l$ ). Отже загальне число нащадків дорівнює  $\xi_1 + \xi_2 + \dots + \xi_i$ .

**Означення 1.** Генератрисою цілочисельної невід’ємної випадкової величини  $\xi$ , для якої  $P\{\xi = k\} = p_k$ ,  $k = 0, 1, 2, \dots$  називається функція

$$g(s) = \sum_{k=0}^{\infty} p_k s^k.$$

Оскільки  $p_k \geq 0$  і  $\sum_{k=0}^{\infty} p_k = 1$ , то функція  $g(s)$  визначена принаймні для тих  $s$ , що  $|s| \leq 1$  ( $s$  — комплексна змінна) і нескінченно диференційовна при  $|s| < 1$ .

*Зуваження 11.* Генератриса суми незалежних цілочисельних невід’ємних випадкових величин дорівнює добутку генератрис окремих доданків.

Отже, генератриса суми  $\xi_1 + \xi_2 + \dots + \xi_i$  дорівнює  $[g(s)]^i$ , де  $g$  — генератриса в.в.  $\xi$ . Тому  $P_{ij}$  є коефіцієнтом при  $j$ -ій степені в степеневому розкладі функції  $[g(s)]^i$ .

## 5.2. Співвідношення для генератрис, що описують гіллястий процес

Отримаємо деякі співвідношення для генератрис величин  $X_n$ . Припустимо, що початкова популяція складається з одного індивідууму, тобто  $X_0 = 1$ . Очевидно, для будь-якого  $n = 0, 1, 2, \dots$  можна записати

$$X_{n+1} = \sum_{r=1}^{X_n} \xi_r,$$

де  $\xi_r$  ( $r \geq 1$ ) — незалежні однаково розподілені випадкові величини з розподілом

$$P\{\xi_r = k\} = p_k, \quad k = 0, 1, 2, \dots, \quad \sum_{k=0}^{\infty} p_k = 1.$$

Введемо генератриси

$$\varphi(s) = \sum_{k=0}^{\infty} p_k s^k$$

та

$$\varphi_n(s) = \sum_{k=0}^{\infty} P\{X_n = k\} s^k, \quad n = 0, 1, 2, \dots$$

Зрозуміло, що

$$\varphi_0(s) = s, \quad \varphi_1(s) = \varphi(s).$$

Далі,

$$\begin{aligned} \varphi_{n+1}(s) &= \sum_{k=0}^{\infty} P\{X_{n+1} = k\} s^k \\ &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P\{X_{n+1} = k | X_n = j\} P\{X_n = j\} s^k \\ &= \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} P\{X_n = j\} P\{\xi_1 + \dots + \xi_j = k\} \\ &= \sum_{j=0}^{\infty} P\{X_n = j\} \sum_{k=0}^{\infty} P\{\xi_1 + \dots + \xi_j = k\} s^k. \end{aligned}$$

Оскільки  $\xi_r$  ( $r = 1, 2, \dots, j$ ) — незалежні однаково розподілені випадкові величини зі спільною генератрисою  $\varphi(s)$ , сума  $\xi_1 + \dots + \xi_j$  має генератрису  $[\varphi(s)]^j$ . Таким чином

$$\varphi_{n+1}(s) = \sum_{j=0}^{\infty} P\{X_n = j\} [\varphi(s)]^j.$$

Але права частина цієї рівності дорівнює генератрисі  $\varphi_n(\cdot)$  з аргументом  $\varphi(s)$ . Таким чином,

$$\varphi_{n+1}(s) = \varphi_n(\varphi(s)). \quad (5.3)$$

Ітеруючи це співвідношення, отримаємо

$$\begin{aligned} \varphi_{n+1}(s) &= \varphi_n(\varphi(s)) = \varphi_{n-1}(\varphi(\varphi(s))) = \varphi_{n-1}(\varphi_2(s)) \\ &= \varphi_{n-2}(\varphi_2(\varphi(s))) = \varphi_{n-2}(\varphi_3(s)). \end{aligned}$$

За індукцією отримуємо, що для будь-якого  $k = 0, 1, \dots, n$

$$\varphi_{n+1}(s) = \varphi_{n-k}(\varphi_{k+1}(s));$$

зокрема, при  $k = n - 1$

$$\varphi_{n+1}(s) = \varphi(\varphi_n(s)). \quad (5.4)$$

Якщо замість умови  $X_0 = 1$  розглянути більш загальну умову  $X_0 = i_0$ , де  $i_0$  — довільне стале число, то

$$\varphi_0(s) = s^{i_0}, \quad \varphi_1(s) = [\varphi(s)]^{i_0},$$

оскільки

$$X_1 = \sum_{j=1}^{i_0} \xi_j.$$

Рівність

$$\varphi_{n+1}(s) = \varphi_n(\varphi(s))$$

зберігається, але співвідношення (5.4) не виконується.

За допомогою (5.3) можна тепер обчислити математичне сподівання та дисперсію величини  $X_n$ . Далі скрізь вважатимемо, що  $X_0 = 1$  (якщо не сказано супротивне). Припустимо, що

$$m = E(X_1) \text{ та } \sigma^2 = \sigma^2(X_1) = E(X_1^2) - (EX_1)^2$$

існують та скінчені.

Очевидно,  $EX_n = \varphi'_n(1)$ . Диференціюючи (5.3) та поклавши  $s = 1$ , отримаємо (оскільки  $\varphi(1) = 1$ ), що

$$\varphi'_{n+1}(1) = \varphi'_n(1)\varphi'(1),$$

або за індукцією

$$\varphi'_{n+1}(1) = \varphi'(1)\varphi'_n(1) = [\varphi'(1)]^2\varphi'_{n-1}(1) = [\varphi'(1)]^3\varphi'_{n-2}(1).$$

За індукцією отримаємо

$$\varphi'_{n+1}(1) = [\varphi'(1)]^n\varphi'_1(1) = [\varphi'(1)]^{n+1}.$$

Але  $\varphi'(1) = \varphi'_1(1) = \mathbf{E}X_1 = m$ . Таким чином,

$$\mathbf{E}X_{n+1} = m^{n+1}. \quad (5.5)$$

Щоб знайти  $\sigma^2(X_{n+1})$ , помітимо спочатку, що

$$\varphi''_n(1) = \sum_{k=2}^{\infty} k(k-1)\mathbf{P}\{X_n = k\} = \mathbf{E}X_n^2 - \mathbf{E}X_n = \mathbf{E}X_n^2 - \varphi'_n(1)$$

і звідси

$$\sigma^2(X_n) = \varphi''_n(1) + \varphi'_n(1) - [\varphi'_n(1)]^2.$$

Диференціюючи (5.4) двічі та поклавши  $s = 1$ , отримаємо

$$\varphi''_{n+1}(1) = \varphi''(1)[\varphi'_n(1)]^2 + \varphi'(1)\varphi''_n(1).$$

Оскільки  $\varphi'(1) = m$ , а  $\varphi''(1) = \mathbf{E}X_1^2 - \mathbf{E}X_1 = \sigma^2 + m^2 - m$ , маємо

$$\varphi''_{n+1}(1) = Lm^{2n} + m\varphi''_n(1),$$

де  $L = \sigma^2 + m^2 - m$ .

За індукцією

$$\begin{aligned} \varphi''_{n+1}(1) &= L [m^{2n} + m^{2n-1}] + m^2 \varphi''_{n-1}(1) \\ &= \dots = L [m^{2n} + m^{2n-1} + \dots + m^n]. \end{aligned}$$

Таким чином,

$$\begin{aligned} \sigma^2(X_{n+1}) &= (\sigma^2 + m^2 - m) [m^{2n} + m^{2n-1} + \dots + m^n] + m^{n+1} - m^{2n+2} \\ &= \sigma^2 [m^{2n} + m^{2n-1} + \dots + m^n] = \sigma^2 m^n \frac{m^{n+1} - 1}{m - 1}, \text{ якщо } m \neq 1. \end{aligned}$$

Якщо  $m = 1$ , то

$$\sigma^2(X_{n+1}) = (n+1)\sigma^2.$$

Ми отримали формули  $\mathbf{E}X_n = m^n$  та

$$\sigma^2(X_n) = \begin{cases} \sigma^2 m^{n-1} \frac{m^n - 1}{m - 1}, & \text{якщо } m \neq 1 \\ n\sigma^2, & \text{якщо } m = 1. \end{cases}$$

Таким чином дисперсія, збільшується (зменшується) зі швидкістю прогресії, якщо  $m > 1$  ( $m < 1$ ), та змінюється лінійно, якщо  $m = 1$ .



### 5.3. Ймовірності виродження

Ми хочемо знайти ймовірність того, що популяція виродиться, тобто  $P\{X_n = 0 \text{ для деякого } n\}$ . Очевидно, якщо  $X_n = 0$ , то  $X_k = 0$  при всіх  $k > n$ .

Помітимо спочатку, що виродження ніколи не відбудеться, якщо ймовірність того, що індивідуум не породжує жодного нащадка, дорівнює нулю, тобто якщо  $p_0 = 0$ . Таким чином, при дослідженні ймовірності виродження припустимо, що  $0 < p_0 < 1$ . Нехай

$$q_n = P\{X_n = 0\} = \varphi_n(0).$$

Як було показано,  $\varphi_{n+1}(s) = \varphi(\varphi_n(s))$ . Тому

$$q_{n+1} = \varphi_{n+1}(0) = \varphi(\varphi_n(0)) = \varphi(q_n). \quad (5.6)$$

Оскільки  $\varphi(s)$  — строго зростає (степеневий ряд з невід'ємними коефіцієнтами і  $p_0 < 1$ ) та  $q_1 = \varphi_1(0) = p_0 > 0$ , то  $q_2 = \varphi(q_1) > \varphi(0) = q_1$ . Припустимо, що  $q_n > q_{n-1}$ . Тоді  $q_{n+1} = \varphi(q_n) > \varphi(q_{n-1}) = q_n$ . Цим доведено, що  $q_1, q_2, \dots, q_n, \dots$  — монотонно зростаюча послідовність, обмежена одиницею. Отже існує

$$\pi = \lim_{n \rightarrow \infty} q_n$$

і  $0 < \pi \leq 1$ . Оскільки  $\varphi(s)$  — неперервна при  $0 \leq s \leq 1$ , покладаючи  $n \rightarrow \infty$  в (5.6), отримуємо

$$\pi = \varphi(\pi). \quad (5.7)$$

Оскільки  $q_n$  — ймовірність виродження популяції не пізніше, ніж за  $n$  поколінь, то  $\pi$  — ймовірність виродження популяції і з (5.7) випливає, що  $\pi$  — корінь рівняння

$$\varphi(s) = s. \quad (5.8)$$

Покажемо тепер, що  $\pi$  — найменший додатний корінь рівняння (5.8). Нехай  $s_0$  — додатний корінь рівняння (5.8). Тоді  $q_1 = \varphi(0) < \varphi(s_0) = s_0$ . Припустимо, що  $q_n < s_0$ . Тоді в силу (5.6)  $q_{n+1} = \varphi(q_n) < \varphi(s_0) = s_0$ . Таким чином, за індукцією  $q_n < s_0$  для всіх  $n$ . Звідси випливає, що

$\pi = \lim_{n \rightarrow \infty} q_n \leq s_0$ , тобто  $\pi$  — найменший додатний корінь рівняння (5.8).

Тепер припустимо, що  $p_0 + p_1 < 1$ . Тоді  $\varphi(s)$  — опукла функція при  $0 < s \leq 1$ , оскільки

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2} > 0.$$

Отже графік функції  $\varphi(s)$  може перетинати пряму з нахилом  $45^\circ$ , що йде з початку координат, найбільше в двох точках. Ми знаємо, що  $\varphi(1) = 1$ , і тому перетин обов'язково відбудеться в точці  $(1, 1)$ .

Очевидно, може мати місце один з двох наступних випадків. Якщо  $m = \varphi'(1) > 1$ , то тангенс кута нахилу дотичної до графіку  $\varphi(s)$  в точці  $s = 1$  більше і має місце випадок 1). В цьому випадку  $0 < \pi < 1$ . Якщо  $m = \varphi'(1) \leq 1$ , то тангенс кута нахилу дотичної в точці  $s = 1$  менше або дорівнює 1 і має місце випадок 2). Тоді  $\pi = 1$ . Таким чином ми довели, що ймовірність виродження дорівнює 1, якщо  $m \leq 1$ , і менше 1, якщо  $m > 1$ . Іншими словами, виродження має місце тоді і тільки тоді, коли середнє число нащадків від одного індивідуума не перевищує 1.

Далі, помітимо, що  $\varphi(s) \leq \varphi(\pi) = \pi$  при  $0 \leq s \leq \pi$ . За індукцією маємо  $\varphi_n(s) \leq \pi$  ( $0 \leq s \leq \pi$ ) для всіх  $n$ . Але  $\varphi_n(s) \geq \varphi_n(0) = q_n$ , і, таким чином,  $q_n \leq \varphi_n(s) \leq \pi$ .

Нехай  $n \rightarrow \infty$ . Тоді

$$\lim_{n \rightarrow \infty} \varphi_n(s) = \pi \text{ при } 0 \leq s \leq \pi.$$

У випадку  $m > 1$ , коли  $\pi < s < 1$ , маємо  $\pi < \varphi(s) < s < 1$ . За індукцією

$$\pi < \varphi_n(s) < \varphi_{n-1}(s) < \dots \quad (\pi < s < 1).$$

Звідси

$$\lim_{n \rightarrow \infty} \varphi_n(s) \geq \pi. \quad (5.9)$$

Границя в (5.9) повинна дорівнювати  $\pi$ , оскільки якби  $\lim_{n \rightarrow \infty} \varphi_n(s) = \alpha > \pi$ , то  $\varphi(\alpha) < \alpha$  і вказана в (5.9) збіжність була б неможливою, бо

$$\lim_{n \rightarrow \infty} \varphi_{n+1}(s) = \lim_{n \rightarrow \infty} \varphi(\varphi_n(s)).$$

Таким чином,

$$\lim_{n \rightarrow \infty} \varphi_n(s) = \pi \text{ при } 0 \leq s < 1.$$

З того, що  $\varphi_n(s)$  збігається до сталої  $\pi$  при  $0 \leq s < 1$ , випливає, що в розкладі

$$\varphi_n(s) = \sum_{k=0}^{\infty} P\{X_n = k\} s^k$$

перший коефіцієнт  $P\{X_n = 0\}$  збігається до  $\pi$  при  $n \rightarrow \infty$ , а всі інші коефіцієнти  $P\{X_n = k\}$  збігаються до 0 при  $n \rightarrow \infty$ ,  $k = 1, 2, \dots$

Отже при будь-якому значенні  $m = EX_1 > 1$  ймовірність того, що  $n$ -те покоління буде складатися з будь-якого додатнього скінченного числа індивідуумів, прямує до 0 при  $n \rightarrow \infty$ , в той час, коли ймовірність виродження прямує до  $\pi$ . В цьому випадку ми скажемо, що  $X_n \rightarrow \infty$  при  $n \rightarrow \infty$  з імовірністю  $1 - \pi$ .

Цей результат є також наслідком загальної теорії ланцюгів Маркова, оскільки ланцюг Маркова, який визначається послідовністю  $X_0, X_1, X_2, \dots$ , має єдиний поглинаючий стан  $\{0\}$ , і тому  $\lim_{n \rightarrow \infty} P_{ij}^n = 0$ ,  $1 \leq i, j < \infty$ , оскільки  $i$  та  $j$  є автоматично перехідними станами (бо з них досяжний стан  $\{0\}$ ).

#### 5.4. Приклади

(1) Нехай  $\varphi(s) = p_0 + p_1 s$ ,  $0 < p_0 < 1$ . Відповідний гіллястий процес є процесом чистої загибелі. В будь-який період часу кожен окремий індивідуум гине з імовірністю  $p_0$  і залишається жити з імовірністю  $p_1 = 1 - p_0$ .

(2) Нехай  $\varphi(s) = p_0 + p_2 s^2$  ( $0 < p_0 < 1$ ,  $p_0 + p_1 = 1$ ). Така генератриса відповідає гіллястому процесу, в якому в кожному поколінні окремий індивідуум або гине, або породжує двох нащадків.

(3) Розглянемо приклад, де кожен індивідуум породжує  $N$  або 0 прямих нащадків з імовірностями  $p$  та  $q$  відповідно. Таким чином,  $p_0 = q$ ,

$p_N = p$  і  $p_k = 0$  для  $k \neq 0, N$ . Тоді

$$\varphi(s) = q + ps^N. \quad (5.10)$$

(4) Кожен індивідуум може мати  $k$  нащадків, де  $k$  — випадкова величина з біноміальним розподілом ймовірностей з параметрами  $N$  та  $p$  ( $P\{k = n\} = C_n^k p^n q^{N-n}$ ,  $n = 0, 1, \dots, N$ ). Тоді

$$\varphi(s) = (q + ps)^N. \quad (5.11)$$

(5) В зв'язку з прикладом (б), що описаний на початку розділу, часто припускається, що випадкове число  $k$  ( $k = 0, 1, 2, \dots$ ) прямих “нащадків” мутантного гену має пуасонівський розподіл  $\left( P\{k = n\} = \frac{e^{-\lambda} \lambda^n}{n!}, n = 0, 1, 2, \dots \right)$  з середнім  $\lambda = 1$ . Тоді

$$\varphi(s) = e^{s-1}$$

і  $\pi = 1$  (корінь рівняння  $e^{s-1} = s$ ).

Обґрунтування для вибору саме цього розподілу наступні. В багатьох популяціях виробляється велике число зигот (запліднених яєць) і лише мала частина їх доживає до зрілості. Події, що полягають в заплідненні та виживанні до зрілості, відбуваються за законом незалежних біноміальних випробувань. Число випробувань (тобто число зигот) настільки велике, що реальне число зрілих нащадків відповідає пуасонівському розподілу.

Саме наслідок закону рідких подій виправдовує застосування пуасонівського наближення. Воно здається досить прийнятним в моделі зростання популяції рідкісних мутантних генів. Якщо мутантний ген має деяку сприятливу (або несприятливу) біологічну ознаку, то береється розподіл Пуасона з середнім  $\lambda > 1$  (або  $\lambda < 1$ ). В цьому випадку

$$\varphi(s) = e^{\lambda(s-1)} \quad (5.12)$$

і  $0 < \pi < 1$  тоді й тільки тоді, коли  $\lambda > 1$ .

*Зауваження 12.* Біноміальний закон описує розподіл числа успіхів в послідовності з  $N$  незалежних випробувань, коли ймовірність успіху в

одному випробування дорівнює  $p$ .  $P\{\xi_{N,p} = n\} = C_N^n p^n q^{N-n}$ ,  $q = 1 - p$  — ймовірність невдачі.

Закон Пуассона є граничним випадком біноміального розподілу, коли ймовірність успіху  $p$  дуже мала, але число експериментів  $N$ , що відбуваються за одиницю часу, досить велике, а саме таке, що в процесі  $p \rightarrow 0$  і  $N \rightarrow \infty$  добуток  $Np \rightarrow \lambda$  (до деякої додатньої сталлої). Тому закон Пуассона часто називають законом рідких подій. Нехай  $\xi_0$  — пуассонівська в.в. Тоді

$$\begin{aligned} P\{\xi_0 = n\} &= \lim_{\substack{p \rightarrow 0 \quad N \rightarrow \infty \\ (pN \rightarrow \lambda)}} P\{\xi_{N,p} = n\} = \lim_{\substack{p \rightarrow 0 \quad N \rightarrow \infty \\ (pN \rightarrow \lambda)}} C_N^n p^n (1-p)^{N-n} \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1)\dots(N-n+1)}{n!} \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \\ &= \lim_{N \rightarrow \infty} \frac{\lambda^n}{n!} \left(1 - \frac{\lambda}{N}\right)^N \frac{N-1}{N} \frac{N-2}{N} \dots \frac{N-n+1}{N} \left(1 - \frac{\lambda}{N}\right)^{-n} \\ &= \frac{\lambda^n}{n!} \lim_{N \rightarrow \infty} \left(1 - \frac{\lambda}{N}\right)^N \\ &\quad \cdot \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{n-1}{N}\right) \left(1 - \frac{\lambda}{N}\right)^{-n} \\ &= \frac{\lambda^n}{n!} e^{-\lambda}. \end{aligned}$$

У випадку гетерогенної (неоднорідної) популяції мутантних генів можна припустити, що ймовірнісний розподіл числа нащадків є пуассонівським, але з випадковим середнім. Наприклад, можна уявити велику географічну область, в кожній підобласті якої гіллястий процес характеризується генератрисою пуассонівського розподілу з параметром  $\lambda$ . Припустимо далі, що значення  $\lambda$  змінюється в залежності від підобласті і розподіл значення  $\lambda$  у всій області є гамма-розподілом. Формально постулюється, що ймовірність того, що мутантний ген має рівно  $k$  прямих “нащадків”, дорівнює

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots,$$

де  $\lambda$  — випадкова величина, яка має гамма-розподіл зі щільністю

$$f(\lambda) = \begin{cases} \frac{1}{\Gamma(\alpha)} \left(\frac{q}{p}\right)^\alpha \lambda^{\alpha-1} \exp\left\{-\frac{q}{p}\lambda\right\}, & \lambda \geq 0, \\ 0, & \text{в інших випадках;} \end{cases}$$

тут  $q, p, \alpha$  — додатні сталі,  $q + p = 1$ .

*Зауваження 13.* Це гамма розподіл з параметрами  $\alpha > 0$ ,  $\frac{q}{p} > 0$ .  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$  — гамма-функція.

Якщо усереднити за параметром  $\lambda$ , то отримаємо ймовірність того, що індивідуум має  $k$  нащадків:

$$P\{\xi = k\} = \int_0^\infty P\{\xi = k|\lambda\} f(\lambda) d\lambda.$$

Генератриса дорівнює

$$\begin{aligned} \varphi(s) &= \sum_{k=0}^{\infty} P\{\xi = k\} s^k \\ &= \sum_{k=0}^{\infty} \int_0^\infty \exp\{-\lambda\} \frac{1}{k!} \lambda^k \frac{1}{\Gamma(\alpha)} \left(\frac{q}{p}\right)^\alpha \lambda^{\alpha-1} \exp\left\{-\frac{q}{p}\lambda\right\} s^k d\lambda \\ &= \int_0^\infty \exp\{-\lambda\} \frac{1}{\Gamma(\alpha)} \left(\frac{q}{p}\right)^\alpha \lambda^{\alpha-1} \exp\left\{-\frac{q}{p}\lambda\right\} \left(\sum_{k=0}^{\infty} \frac{1}{k!} (\lambda s)^k\right) d\lambda \\ &= \int_0^\infty \exp\left\{-\left(\frac{q}{p} + 1 - s\right)\lambda\right\} \frac{1}{\Gamma(\alpha)} \left(\frac{q}{p}\right)^\alpha \lambda^{\alpha-1} d\lambda \\ &= \int_0^\infty \exp\left\{-\left(\frac{q}{p} + 1 - s\right)\lambda\right\} \frac{1}{\Gamma(\alpha)} \left(\frac{q}{p} + 1 - s\right)^\alpha \lambda^{\alpha-1} d\lambda \\ &\quad \times \left(\frac{q/p}{q/p + 1 - s}\right)^\alpha = \left(\frac{q/p}{q/p + 1 - s}\right)^\alpha = \left(\frac{q}{1 - ps}\right)^\alpha. \end{aligned}$$

Ми отримали генератрису від'ємного біноміального розподілу з параметрами  $\alpha > 0$  та  $0 < q < 1$ :  $P\{\xi = k\} = C_{\alpha+n-1}^n q^\alpha p^n$ .

## 6. Процеси Маркова, що породжуються гіллястими процесами

Розглянемо спочатку кілька прикладів ланцюгів Маркова, що виникають в генетиці.

### 6.1. Модель Райта

Наступна ідеалізована генетична модель була запропонована С. Райтом для дослідження флуктуації частот генотипів під впливом мутації та відбору. Почнемо з так званої простої гапліодної моделі випадкового відтворення, в якій не враховуються дія процесу мутації та сили відбору. Припустимо, що ми маємо справу з популяцією фіксованого об'єму, яка складається з  $2N$  індивідумів типів  $a$  та  $A$ . Склад наступного покоління визначається результатом  $2N$  випробувань Бернуллі, а саме: якщо популяція батьків складається з  $j$  генів типу  $a$  та  $2N - j$  генів типу  $A$ , то ймовірність появи гену  $a$  або гену  $A$  при кожному випробуванні відповідно дорівнюють

$$p_j = \frac{j}{2N}, \quad q = 1 - \frac{j}{2N},$$

що еквівалентно випадковому вибору з поверненням. Ця процедура описується ланцюгом Маркова  $\{X_n\}$ , де  $X_n$  — число  $a$ -генів в  $n$ -му поколінні популяції. Простором станів ланцюга є набір з  $2N + 1$  чисел  $\{0, 1, 2, \dots, 2N\}$ . Елементи матриці перехідних імовірностей обчислюються у відповідності до біноміального розподілу:

$$P\{X_{n+1} = k | X_n = j\} = P_{jk} = C_{2N}^k p_j^k q_j^{2N-k}, \quad j, k = 0, 1, \dots, 2N. \quad (6.1)$$

Відмітимо, що стани  $0$  та  $2N$  є поглинаючими в тому сенсі, що якщо  $X_n = 0$  або  $2N$ , то  $X_{n+k} = 0$  або  $2N$  відповідно для всіх  $k \geq 0$ .

Більш реалістична модель враховує мутаційний тиск. Припустимо, що перш, ніж буде утворено нове покоління, кожен ген має можливість мутувати, тобто перетворитися в ген іншого типу. Тобто, ми припустимо, що для всіх генів типу  $a$  ймовірність мутації  $a \rightarrow A$  дорівнює  $\alpha_1$ , а

для генів типу  $A$  ймовірність мутації  $A \rightarrow a$  дорівнює  $\alpha_2$ . Як і раніше, будемо вважати, що склад кожного наступного покоління визначається результатом  $2N$  випробувань Бернуллі. В тому випадку, коли батьківська популяція містить  $j$  генів типу  $a$ , відповідні величини  $p_j$  та  $q_j$  мають вигляд:

$$\begin{aligned} p_j &= \frac{j}{2N}(1 - \alpha_1) + \left(1 - \frac{j}{2N}\right) \alpha_2, \\ q_j &= \frac{j}{2N} \alpha_1 + \left(1 - \frac{j}{2N}\right) (1 - \alpha_2). \end{aligned} \tag{6.2}$$

Суть справи полягає в наступному: згідно цієї моделі, спочатку діє мутаційний процес, після чого тип кожного гену наступного покоління визначається випадковим вибором (з поверненням) із батьківської популяції. Ймовірність вибору  $a$ -гену з популяції, що зазнала дію мутаційного процесу, дорівнює числу  $a$ -генів в цій популяції, помноженому на  $\frac{1}{2N}$ ; тому середня ймовірність (усереднена по можливим мутаціям) є добутком середнього числа  $a$ -генів після мутації на  $\frac{1}{2N}$ . Але це число, очевидно, дорівнює  $j(1 - \alpha_1) + (2N - j)\alpha_2$ , що й призводить до (6.2).

Перехідні ймовірності відповідного ланцюга Маркова обчислюються за формулами (6.1) з використанням величин  $p_j$  та  $q_j$  з (6.2). Якщо  $\alpha_1, \alpha_2 > 0$ , то система на фіксується в жодному зі станів. При  $n \rightarrow \infty$  розподіл імовірностей в.в.  $X_n$  прямує до розподілу деякої випадкової величини  $\xi$ :

$$P\{\xi = k\} = \pi_k, \quad k = 0, 1, \dots, 2N, \quad \sum_{k=0}^{2N} \pi_k = 1, \quad \pi_k > 0.$$

Цей розподіл називається стаціонарним розподілом частот генотипу в популяції.

Розглянемо просту модель випадкового схрещування та введемо поняття сили відбору, що діє, скажімо, на користь  $a$ -генотипу. Припустимо, що ми хочемо визначити селективну перевагу  $a$ -гену над  $A$ -геном так, щоб математичні сподівання відносного числа нащадків були пропорційні  $1 + s$  та  $1$  відповідно. Замінімо величини  $p_j = \frac{j}{2N}$  та



$$q_j = 1 - \frac{j}{2N} \text{ на}$$

$$p_j = \frac{(1+s)j}{2N+sj} \text{ та } q_j = 1 - p_j$$

та визначимо склад наступного покоління, як і раніше, за допомогою випробувань Бернуллі. Якщо батьківська популяція містила  $j$  генів типу  $a$ , то в наступному поколінні середні значення числа генів типу  $a$  та числа генів типу  $A$  дорівнюють відповідно

$$2N \frac{(1+s)j}{2N+sj} \text{ та } 2N \frac{2N-j}{2N+sj}.$$

Відношення середнього числа  $a$ -генів до середнього числа  $A$ -генів в  $(n+1)$ -му поколінні дорівнює

$$\frac{1+s}{1} \cdot \frac{j}{2N-j} = \left( \frac{1+s}{1} \right) \left( \frac{\text{число } a\text{-генів в } n\text{-му поколінні}}{\text{число } A\text{-генів в } n\text{-му поколінні}} \right).$$

Це співвідношення пояснює сутність дії відбору.

## 6.2. Модель полісомного спадкування

Ген складається з деякого числа, скажімо,  $N$  елементарних одиниць. Пред тим, як клітина, що містить ген, починає ділитися, кожна з цих одиниць подвоюється, і кожна з двох “дочірніх” клітин отримує ген, що складається з тієї ж кількості елементарних одиниць, що й ген “материнської” клітини. Одна або декілька одиниць можуть бути в мутаційній формі. Припускається, що в процесі подвоєння гену кожна мутантна одиниця перетворюється на дві мутантні одиниці, а кожна немутантна одиниця — на дві немутантні. Крім того, ми припускаємо, що після подвоєння ці одиниці розподіляються між двома новими генами випадковим чином, так як би вони витягалися з урни. Ми будемо казати, що ген знаходиться в стані  $i$ , якщо він містить  $i$  мутантних та  $N-i$  немутантних одиниць. Перехідні ймовірності ланцюга обчислюються за формулою

$$P_{ij} = \frac{C_{2i}^j C_{2N-2i}^{N-j}}{C_{2N}^N}. \quad (6.3)$$

Цей вираз для  $P_{ij}$  отримано наступним чином. Припустимо, що батьківський ген знаходиться в стані  $i$ ; тоді після подвоєння ми отримаємо набір з  $2i$  мутантних та  $2N - 2i$  немутантних одиниць. Для того, щоб утворити дочірній ген, з цього набору беруться  $N$  довільних одиниць. Тоді ймовірність того, що дочірній ген опиниться в стані  $j$ , визначається формулою (6.3).

Стани  $j = 1, 2, \dots, N - 1$  називаються змішаними, а стани  $0$  та  $N$  — чистими. Стан  $N$  цікавий тим, що ген, всі елементарні одиниці якого мутантні, може принести смерть своєму власнику, в той час як ген, що знаходиться в стані  $0$ , при діленні не породжує мутантних генів.

### 6.3. Процеси Маркова, що породжуються прямим добутком гіллястих процесів

Існує тісний зв'язок між теорією гіллястих процесів та деякими ймовірнісними моделями частот, що використовуються в генетиці та екології популяцій. Ми введемо в розгляд клас скінчених ланцюгів Маркова спеціального виду, який включає в себе як часткові випадки ланцюги, що розглядалися перед цим в якості прикладів застосування до генетики.

Нехай  $f(s) \sum_{n=0}^{\infty} a_n s^n$ ,  $\sum_{n=0}^{\infty} a_n = 1$ ,  $a_n \geq 0$ . Тоді марківський гіллястий процес визначається як ланцюг Маркова, для якого елемент матриці перехідних імовірностей  $P_{ij}$  визначається як коефіцієнт при  $s^j$  в розкладі  $f^i(s)$ . Імовірнісний розподіл, що визначається послідовністю  $\{a_i\}$ , називається розподілом числа нащадків для одного індивідууму, а  $f^i(s)$  — імовірнісною генератрисою для числа незалежно породжених нащадків  $i$  індивідуумів. Аналогічно процес  $\{X_n\}$  є гіллястим, якщо  $X_{n+1} = \sum_{k=1}^i \xi_k = j$ , де  $X_n = i$  — число індивідуумів в поточному поколінні,  $\xi_k$  — число нащадків  $k$ -го індивідууму в поточному поколінні, а  $j$  — число індивідуумів в наступному поколінні. Випадкові величини  $\xi_1, \xi_2, \dots, \xi_i$  є незалежно розподіленими із загальною ймовірнісною

генератрисою  $f(s)$ .

Розглянемо тепер дві популяції типів  $A$  та  $a$ , кожна з яких незалежним чином розмножується, як гіллястий процес. Це — двовимірний процес, що розгортається як послідовність пар випадкових величин  $Z_n = (X_n; Y_n)$ , де  $X_n$  та  $Y_n$  — відповідно числа типів  $A$  та  $a$  в  $n$ -му поколінні. Стан процесу описується парою  $(i, j)$ ,  $i \geq 0$ ,  $j \geq 0$ , де  $i$  та  $j$  — поточні значення числа типів  $A$  та  $a$  відповідно. В такому формулюванні компоненти  $X_n$  та  $Y_n$  утворюють незалежні гіллясті процеси. Тому можна назвати  $Z_n$  *прямим добутком гіллястих процесів*  $\{X_n\}$  та  $\{Y_n\}$ .

Нехай  $f(z)$  — генератриса розподілу числа нащадків індивідуумів типу  $A$ , а  $g(w)$  — генератриса розподілу числа нащадків індивідуумів типу  $a$ . Якщо індивідууми розвиваються незалежно один від одного, то компоненти матриці перехідних імовірностей обчислюються за формулою

$$\begin{aligned} P\{X(1) = j_1, Y(1) = j_2 | X(0) = i_1, Y(0) = i_2\} &= P_{(i_1, i_2)(j_1, j_2)} \\ &= \text{коефіцієнт при } z^{j_1} w^{j_2} \text{ в розкладі } f^{i_1}(z)g^{i_2}(w). \end{aligned}$$

Можна сформулювати марківську модель частот, пов'язану з гіллястим процесом для двох типів індивідуумів, в якій розмір популяції сталий в часі. Маючи це на увазі, знайдемо спочатку ймовірність того, що в поколінні нащадків є  $j$  індивідуумів типу  $A$  за умови, що загальне число нащадків дорівнює  $M$  та поточне покоління складається з  $i$  індивідуумів типу  $A$  та  $N - i$  індивідуумів типу  $a$ :

$$\begin{aligned} P\{X(1) = j | X(0) = i, Y(0) = N - i, X(1) + Y(1) = M\} \\ &= \frac{P\{X(1) = j, Y(1) = M - j | X(0) = i, Y(0) = N - i\}}{P\{X(1) + Y(1) = M | X(0) = i, Y(0) = N - i\}} \\ &= \frac{\text{коефіцієнт при } z^j w^{M-j} \text{ в розкладі } f^i(z)g^{N-i}(w)}{\text{коефіцієнт при } t^M \text{ в розкладі } f^i(t)g^{N-i}(t)}. \end{aligned} \quad (6.4)$$

Тут використано той факт, що  $f^i(t)g^{N-i}(t)$  — генератриса загального числа нащадків, що не розрізняються за типами. Якщо  $N = M$ , то можна вважати, що вираз (6.4) визначає матрицю перехідних імовірностей

$P_{ij}(N)$  скінченного ланцюга Маркова з простором станів  $\{0, 1, 2, \dots, N\}$ .

Відношення номеру стану до числа  $N$  можна інтерпретувати як частоту типу  $A$  у всій популяції. Отже, побудований ланцюг Маркова є дискретною ймовірнісною моделлю, яка описує зміну частот типів в популяції сталого розміру, що складається з двох типів індивідуумів. Зміни частоти в кожному з поколінь визначається виглядом матриці перехідних імовірностей ланцюга Маркова.

Розмір популяції  $N$  є параметром, який часто для спрощення буде опускається. Таким чином,

$$P_{ij} = P\{X(1) = j | X(0) = i, Y(0) = N - i, X(1) + Y(1) = N\} \\ = \frac{\text{коефіцієнт при } z^j w^{N-j} \text{ в розкладі } f^i(z)g^{N-i}(w)}{\text{коефіцієнт при } t^N \text{ в розкладі } f^i(t)g^{N-i}(t)}. \quad (6.5)$$

Ланцюг Маркова, що характеризується матрицею перехідних імовірностей (6.5), іноді називають ланцюгом Маркова, що породжений прямим добутком марківських гіллястих процесів.

Ми проілюструємо можливості даного класу ланцюгів Маркова на трьох найпростіших прикладах та вкажемо їх місце в прикладних задачах.

(1) В якості першого прикладу розглянемо прямий добуток пуасонівських гіллястих процесів, коли

$$f(z) = e^{\lambda(z-1)}, \quad g(w) = e^{\mu(w-1)}, \quad \lambda, \mu > 0.$$

При цьому

$$P_{ij} = \frac{\text{коефіцієнт при } z^j \text{ в розкладі } e^{i\lambda(z-1)}}{\text{коефіцієнт при } t^N \text{ в розкладі } e^{(i\lambda+(N-i)\mu)(t-1)}} \\ \times [\text{коефіцієнт при } w^{N-j} \text{ в розкладі } e^{(N-i)\mu(w-1)}] \\ = \frac{e^{-i\lambda} \left[ \frac{(i\lambda)^j}{j!} \right] \left\{ e^{-(N-i)\mu} \frac{1}{(N-j)!} [(N-i)\mu]^{N-j} \right\}}{e^{-(i\lambda+(N-i)\mu)} \left\{ [i\lambda + (N-i)\mu]^N \frac{1}{N!} \right\}} \\ = C_N^j \left[ \frac{i\lambda}{i\lambda + (N-i)\mu} \right]^j \left[ \frac{(N-i)\mu}{i\lambda + (N-i)\mu} \right]^{N-j}, \quad i, j = 0, 1, 2, \dots, N. \quad (6.6)$$

Якщо  $\lambda = \mu$ , то отримуємо вираз

$$P_{ij} = C_N^j \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}, \quad i, j = 0, 1, \dots, N, \quad (6.7)$$

що не залежить від  $\lambda$ . Це в точності співпадає з ймовірнісною моделлю біноміального вибору, запропонованою С.Райтом для вивчення флуктуації частот гену в гаплоїдній популяції під впливом випадкового дрейфу. Якщо  $\lambda \neq \mu$ , то (6.6) зводиться до моделі відбору Райта, в якій  $\frac{\lambda}{\mu} = 1\sigma$ , а  $\sigma$  — коефіцієнт відбору, що виражає перевагу індивідуумів типу  $A$  над індивідуумами типу  $a$ .

В загальному випадку, що відповідає формулі (6.4), ми скажемо, що відбір має місце, якщо  $f(z) \neq g(z)$ , тобто якщо ймовірнісні генератриси числа нащадків двох типів різні.

(2) Нехай  $f(z) = g(z) = (q + pz)^2$ ,  $q = 1 - p$ ,  $0 < p < 1$ . Породжений ланцюг Маркова (6.5) в цьому випадку набуде вигляду

$$P_{ij} = [C_{2N}^N]^{-1} C_{2i}^j C_{2(N-i)}^{N-j}, \quad i, j = 0, 1, 2, \dots, N. \quad (6.8)$$

Інтерпретація цього ланцюга Маркова як моделі полісомного спадкування дана в другому прикладі.

(3) Нехай

$$f(z) = \frac{q^\alpha}{(1 - pz)^\alpha}, \quad g(z) = \frac{q^\beta}{(1 - pz)^\beta}, \quad q = 1 - p, \quad 0 < p < 1, \quad \alpha, \beta > 0.$$

Тоді, використовуючи розклад

$$\frac{1}{(1 - pz)^\gamma} = \sum_{n=0}^{\infty} C_{\gamma+n-1}^n p^n z^n,$$

ми бачимо, що породжений ланцюг Маркова має наступні перехідні ймовірності

$$P_{ij} = \left[ C_{i\alpha+(N-i)\beta+N-1}^N \right]^{-1} C_{i\alpha+j-1}^j C_{(N-i)\beta+N-i-1}^{N-j}, \quad i, j = 0, 1, \dots, N. \quad (6.9)$$

Цей ланцюг Маркова може інтерпретуватися як модель росту гетерозиготної популяції (приклад (5) гіллястих процесів з генератрисою

$$\varphi(s) = \left( \frac{q}{1 - ps} \right)^\alpha.$$

## Література

- [1] *Карлин, С.* Основы теории случайных процессов. – М.: Мир, 1971. – 536 с.
- [2] *Durrett, Richard.* Probability models for DNA sequence evolution. – Springer, 2008. – 431 p.
- [3] *Berestycki, Nathanaël.* Recent progress in coalescent theory. – *Ensaio Matemáticos* 16 (2009). – P. 1–193.