

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО"

Моделювання систем

Навчальний посібник

*Рекомендовано Вченою радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеню бакалавра
за освітньою програмою "Комп'ютерні технології в біології та медицині"
спеціальності "Комп'ютерні науки"*

Укладачі: **К. Х. Зеленський, Є. А. Настенко, В. А. Павлов**

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Рецензенти:

*О.М. Тачиніна, д.т.н., професор, професор кафедри авіоніки
Національного авіаційного університету*

*А. М. Сільвестров, д.т.н., професор
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
КПІ ім. Ігоря Сікорського*

Відповідальний редактор

*О. К. Городецька, к.т.н., доцент
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"*

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 2 від 30.08.2022 р.) за поданням Вченої ради факультету біомедичної інженерії (протокол № 1 від 01.08.2022 р.)

У навчальному посібнику "Моделювання систем" викладено загальні питання моделювання складних статичних та динамічних систем, підходи до отримання псевдовипадкових чисел як підґрунтя для вирішення задач статистичного моделювання систем. Викладено питання планування багатофакторних експериментів для отримання статичних моделей систем. Викладено методи моделювання динамічних систем за допомогою динамічного програмування, а також адаптивних динамічних систем. Викладання теоретичного матеріалу ілюструється численними прикладами їхнього застосування.

Реєстр. № НП 22/23-126. Обсяг 33 авт. арк.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів і розповсюджувачів видавничої продукції ДК
№ 5354 від 25.05.2017 р.

© КПІ ім. Ігоря Сікорського, 2022

Зміст

1 Загальні питання моделювання систем	7
1.1 Задачі та методи моделювання	7
1.1.1 Приклади математичних моделей	7
1.1.2 Класифікація і методи дослідження математичних моделей	9
1.1.3 Основні властивості математичних моделей	13
1.1.4 Загальна схема математичного моделювання	14
1.1.5 Поняття моделі та способи побудови моделей	17
1.1.6 Методи і процес моделювання	19
1.1.7 Приклади розв'язання задач	22
1.2 Контрольні запитання	24
2 Статистичне моделювання систем	25
2.1 Вступ	25
2.1.1 Граничні теореми теорії ймовірностей	25
2.2 Комп'ютерне моделювання ПВЧ	26
2.2.1 Ідентифікація закону управління	27
2.2.2 Розподіли ймовірностей випадкових величин	28
2.2.3 Апроксимація функціональної залежності	40
2.2.4 Кореляційно-регресійний аналіз функціональної залежності	41
2.2.5 Апроксимація залежності за методом хі-квадрат	43
2.2.6 Приклади розв'язання задач	43
2.3 Генератори ПВЧ	46
2.3.1 Обчислення у полях Галуа	46
2.3.2 Вихор Мерсенна	51
2.4 Створення випадкового числа з біоелектричних та фізичних сигналів	54
2.4.1 Методи генерації випадкових чисел	55
2.4.2 Біоелектричні та фізичні сигнали	56
2.4.3 Генерація особистих ідентифікаторів	57
2.4.4 Завдання для самостійної роботи	59
2.5 Контрольні запитання	61
3 Теорія масового обслуговування	62
3.1 Загальні засади теорії	62
3.2 Методи дослідження СМО	66
3.2.1 Диференційний метод	66
3.2.2 Робота телефонного комутатора	69
3.2.3 Метод евклідових ланцюгів Маркова	73
3.2.4 СМО з очікуванням	75
3.3 Контрольні запитання	76
4 Планування експериментів	79
4.1 Вступ	79
4.2 Характеристика основних задач і методів багатофакторного експерименту	80
4.2.1 Вступ	80
4.2.2 Вибір факторів	80
4.2.3 Вибір основного рівня та інтервалів варіювання	80
4.2.4 Задачі класифікації	81
4.3 План експерименту	82
4.4 Плани першого порядку	84
4.5 Планування експерименту для нелінійних моделей	85

4.5.1 Нелінійний регресійний аналіз.	85
4.5.2 Числові методи пошуку МНК-оцінок.	87
4.5.3 Загальна характеристика методів і задач регресійного аналізу	88
4.6 Плани для квадратичних моделей.	96
4.7 Композиційні ортогональні плани.	100
4.7.1 Повний факторний експеримент (ПФЕ)	100
4.7.2 ЦКП.	101
4.7.3 Ротатбельні центральні композиційні плани.	103
4.8 Побудова канонічної моделі.	106
4.9 Послідовне планування.	109
4.9.1 Лінійна регресія.	111
4.9.2 Метод Йетса	115
4.10 Багаторівневі і несимметричні факторні плани.	116
4.10.1 Загальні поняття про факторні плани та моделі.	117
4.10.2 Багаторівневі факторні плани.	119
4.10.3 Вибір оптимальних планів у практичних застосуваннях	123
4.11 Планування екстремального експерименту.	124
4.11.1 Збіжність псевдоградієнтних алгоритмів	125
4.11.2 Планування екстремального експерименту за наявності обмежень	126
4.12 Симплексний метод оптимізації.	132
4.12.1 Симплексний метод.	140
4.12.2 Симплексний метод оптимального планування експерименту.	142
4.13 Контрольні запитання	145
5 Метод групового урахування аргументів	146
5.1 Вступ.	146
5.2 Постановка задачі.	146
5.3 Багаторядний МГУА.	148
5.3.1 Критерії регулярності.	149
5.3.2 Критерій незміщеності.	150
5.3.3 Критерій балансу змінних.	152
5.3.4 Алгоритм розділення початкової вибірки даних	153
5.4 Опис індуктивного алгоритму.	153
5.4.1 Зовнішні критерії.	155
5.5 Комбінаторний алгоритм МГУА	156
5.6 Контрольні запитання.	159
6 Динамічне програмування.	160
6.1 Дискретні керуючі систем.	161
6.1.1 Адитивні функції і задачі.	162
6.1.2 Загальна схема динамічного програмування.	163
6.1.3 Числова реалізація.	165
6.1.4 Приклади.	167
6.2 Метод ДП для неперервних систем	174
6.2.1 Автономна система	174
6.2.2 Неавтономна система	176
6.3 Контрольні запитання.	177
Моделі біологічних та медичних процесів	
7 Математичне моделювання імунних систем.	178
7.1 Вступ.	178
7.1.1 Математична модель протиінфекційного захисту.	179

7.2 Керування імунологічними реакціями.	182
7.3 Математичне моделювання лікування інфекційних хворих вакцинацією.	185
7.3.1 Огляд математичних моделей в імунології.	185
7.3.2 Побудова моделі.	188
7.3.3 Система рівнянь блоку органу-мішені.	188
7.3.4 Вакцинація.	191
7.4 Регуляція протівірусної відповіді.	199
7.4.1 Математична модель.	199
7.5 Моделі динаміки гепатиту С.	204
7.5.1 Вступ.	204
7.5.2 Біологічна основа: цикл реплікації вірусного геному.	205
7.5.3 Динаміка гепатиту С з оптимальним лікуванням.	206
7.5.4 Моделі ПДЕ.	210
7.5.5 Розширені моделі ПДЕ.	213
7.5.6 Додаток.	215
7.6 Математичні моделі ВІЛ-інфекції.	218
7.6.1 Вступ.	218
7.6.2 Внутрішньоклітинна ВІЛ-модель із затримкою.	218
7.6.3 Оптимальне управління ВІЛ-моделі з затримками.	219
7.6.4 Оптимальне управління моделі ВІЛ з лінійною функцією росту антитіл.	223
7.7 Бібліографія.	228
8 Моделі ракових пухлин.	231
8.1. Загальні засади лікування ракових пухлин.	231
8.2 Оптимальне управління імунотерапією пухлин.	234
8.2.1 Математична модель	234
8.2.2 Вибір параметрів.	238
8.2.3 Приклади числової рівноваги.	238
8.2.4 Оптимальне управління.	240
8.2.5 Числові результати застосування квадратичного програмування.	241
8.3 Моделювання виникнення і росту пухлин: карцинома.	242
8.3.1 Вступ.	242
8.3.2 Загальні особливості математичних моделей ракових пухлин.	242
8.3.3 Гліома.	247
9 Лікування гліоми	250
9.1 Вступ	250
9.2 Віротерапія гліоми.	252
9.2.1 Математична модель лікування гліоми віротерапією.	252
9.2.2 Результати.	255
9.3 Вплив CNN1 і макрофагів на гліому.	258
9.3.1 Математична модель.	258
9.3.2 Оцінка параметрів і вміст макрофагів.	262
9.3.3 Числові результати.	263
9.3.4 Додаток.	265
10 Моделювання інвазії пухлин.	270
10.1 Компоненти системи uPA та їх функції.	270
10.1.1 Активатор плазміногену урокінази.	270
10.1.2 Рецептор активатора плазміногену урокінази.	270
10.1.3 Вітронектин.	270
10.1.4 Інгібітор активатора плазміногену урокінази-1.	271

10.1.5 Плазмін.	271
10.2 Математичне моделювання.	271
10.2.1 Макроскопічна динаміка.	271
10.2.2 Мікроскопічна динаміка.	272
10.2.3 Макроскопічне переміщення межі пухлини.	274
10.3 Результати моделювання.	275
10.3.1 Початковий стан ЕСМ.	277
10.3.2 Коефіцієнт дифузії ракових клітин.	277
10.3.3 Швидкість розповсюдження ЕСМ і швидкість деградації.	278
11 Комбінована терапія меланоми.	282
11.1 Вступ.	282
11.2 Математична модель.	283
11.3 Додаток. Оцінка параметрів моделі.	291
11.4 Бібліографія.	293

Розділ 1. Загальні питання моделювання систем

Метою вивчення дисципліни "Моделювання систем" студентами напрямів "Прикладна математика" та "Комп'ютерні науки" є надання їм початкових знань і навичок, необхідних для побудови і дослідження математичних моделей складних природних, технічних, екологічних, біологічних та організаційних систем і процесів, навчання студентів основним положенням, загальним підходам та методам математичного моделювання.

Передбачається, що студенти мають підготовку з математичного аналізу, алгебри, аналітичної геометрії, теорії диференціальних рівнянь та обчислювальних методів у межах, встановлених вимогами відповідних освітньо-професійних програм. Крім того, передбачається, що студенти знайомі з основами фізики в рамках програми нетехнічних спеціальностей вищих навчальних закладів.

1.1 Задачі та методи моделювання

Математичне моделювання є одним із основних сучасних методів дослідження. Загалом під моделюванням мається на увазі процес дослідження реальної системи, який містить побудову моделі, її дослідження та перенесення отриманих результатів на систему, що моделюється. Модель можна визначити як об'єкт, що у деяких відношеннях збігається із прототипом і є засіб опису, пояснення та/або прогнозування його поведінки. Під математичною моделлю реальної системи (процесу) мається на увазі сукупність співвідношень, які визначають характеристики станів системи залежно від її параметрів, зовнішніх умов (вхідних сигналів, збурень), початкових умов та часу.

1.1.1 Приклади математичних моделей

Рух матеріальної точки під дією постійної сили F можна описати рівнянням:

$$r(t) = r_0 + v_0 t + \frac{Ft^2}{2m}, \quad (1)$$

де $r(t)$ -- радіус-вектор точки у момент часу t , r_0 -- радіус-вектор точки у початковий момент часу, v_0 -- початкова швидкість, m -- маса. Це рівняння можна розглядати як найпростішу математичну модель руху.

У складніших випадках, коли сила не є стала, модель (1) може бути перетворена різними способами. Зокрема, якщо відома величина сили для кожного моменту часу, математичну модель руху можна записати у вигляді системи, що складається із диференційного рівняння другого порядку і двох початкових умов:

$$\frac{d^2 r}{dt^2} = \frac{F}{m}; \quad r(0) = r_0; \quad v(0) = v_0. \quad (2)$$

Якщо для кожного моменту часу відомий радіус-вектор точки, математичну модель руху можна подати у вигляді степеневого ряду Тейлора:

$$r(t) = r_0 + t \frac{dr}{dt} \Big|_{t=0} + \frac{t^2}{2!} \frac{d^2 r}{dt^2} \Big|_{t=0} + \frac{t^3}{3!} \frac{d^3 r}{dt^3} \Big|_{t=0} + \dots + \frac{t^n}{n!} \frac{d^n r}{dt^n} \Big|_{t=0}. \quad (3)$$

У багатьох випадках математичні моделі складних процесів подаються у вигляді систем диференціальних рівнянь. Зокрема, Г.І. Марчук запропонував таку модель взаємодії

імунної системи організму людини з антигеном, [29]:

$$\begin{cases} \frac{dV}{dt} = (\beta - \gamma F)V; \\ \frac{dC}{dt} = \xi Q(t - \tau) - \mu_C(C - C_0); \\ \frac{dF}{dt} = \rho C - (\mu_f + \eta \gamma V)F; \\ \frac{dM}{dt} = \sigma V - \mu_M M. \end{cases} \quad (4)$$

Тут: V -- кількість вірусів в організмі; C -- кількість плазматичних клітин, що виробляють антитіла; F -- кількість антитіл; M -- маса тканини, пошкодженої вірусами; β -- коефіцієнт розмноження вірусів, тобто кількість нових вірусів, яку створює кожен з них у середньому за одиницю часу; γ -- коефіцієнт, що відображає ймовірність нейтралізації вірусу при його зустрічі із антигеном ($\gamma = \text{const}$); ξ -- коефіцієнт, що відображає ймовірність збудження реакції створення каскаду плазматичних клітин одним антигеном; $Q(t - \tau)$ -- швидкість утворення плазматичних клітин; τ -- інкубаційний період, тобто час, потрібний для організації каскаду клітин і початку масового синтезу антитіл; μ_C -- коефіцієнт, зворотно пропорційний часу життя плазматичних клітин; C_0 -- імунологічний рівень плазматичних клітин в організмі; ρ -- коефіцієнт відтворення антитіл за одиницю часу з розрахунку на одну клітину; μ_f -- коефіцієнт, зворотно пропорційний середньому часу розпаду антитіл; η -- коефіцієнт, що відображає ймовірність розпаду антитіла при нейтралізації ним вірусу; σ -- коефіцієнт, що відображає ступінь пошкодження органу одним вірусом за одиницю часу; μ_M -- коефіцієнт, зворотно пропорційний часу відновлення пошкодженої частини органу в e разів.

Рівняння системи (4) відображають відповідно швидкість зростання кількості вірусів, плазматичних клітин, антитіл та маси пошкодженої тканини в організмі. При цьому позитивні доданки відповідають їх утворенню, а негативні -- розпаду.

Операторне рівняння (рівняння Шредингера)

$$-\frac{\hbar}{i} \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + U(x, y, z, t) \psi, \quad (5)$$

де $\hbar$ -- стала Планка, ψ -- хвильова функція, t -- час, x, y, z -- координати частинки, m --

її маса, U -- потенційна енергія, $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, можна розглядати як математичну

модель, що описує зміну стану квантової частинки за часом.

Математична модель електростатичного поля, що відображає теорему Гауса, записується як інтегральне рівняння

$$\oint (E dS) = 4\pi q, \quad (6)$$

де E -- напруженість електричного поля, dS -- елемент поверхні. Ліва частина рівняння (6) описує потік вектора напруженості електричного поля через замкнуту поверхню, всередині якої міститься електричний заряд.

Динамічна модель міжгалузевого балансу Леонтьєва має вигляд матричного рівняння

$$y(t) = Ay(t) + By(t) + c(t), \quad (7)$$

де $x(t)$ -- вектор-стовпець обсягів виробництва, $\dot{x}(t)$ -- вектор-стовпець абсолютних швидкостей приросту виробництва (похідних обсягів виробництва за часом), A -- матриця коефіцієнтів прямих матеріальних витрат, B -- матриця коефіцієнтів капіталомісткості приростів виробництва (витрати виробничого накопичення на одиницю приросту відповідних видів продукції), $c(t)$ -- вектор-стовпець споживання.

1.1.2 Класифікація і методи дослідження математичних моделей

Типовими завданнями моделювання можуть бути пошук оптимальних чи наближених до оптимальних рішень; визначення властивостей системи; встановлення взаємозв'язків між її елементами або характеристиками, а також між характеристиками системи та зовнішнього середовища тощо.

Складні системи можна охарактеризувати функціями, які вони виконують, структурою, а також поведінкою у часі. Відповідно розрізняють функціональні, структурні, інформаційні та поведінкові (подійні) моделі систем. Функціональна модель системи описує сукупність функцій, що їх виконує система (сукупність процесів, які відбуваються у ній), характеризує склад та взаємозв'язки її функціональних підсистем. Структурна модель відображає побудову системи; інформаційна -- відношення між елементами системи, а також між системою і зовнішнім середовищем. Остання будується у вигляді структур даних, що характеризують елементи системи, зовнішнє середовище та взаємозв'язки між ними. Інформаційна модель також може мати вигляд рівнянь регресії або кореляційних рівнянь, які відображають зв'язок між даними, статистичного опису сукупності даних, порівняльних статистичних характеристик наборів даних тощо. Поведінкова модель відображає динаміку функціонування системи, зміни її станів, події, що відбуваються у ній, тощо.

Розглянемо детальніше цю класифікацію на прикладі такої системи, як тверде тіло. Функціональні моделі мають відображати процеси, що відбуваються у системі. Для твердого тіла це можуть бути моделі, які описують теплопровідність, електропровідність, дифузію, коливання кристалічної решітки, поглинання та розсіювання світла тощо. Структурні моделі відображають побудову твердого тіла. Для ідеального кристалу таку модель можна сформулювати як сукупність елементів симетрії кристалічної решітки, як геометричний опис елементарної решітки кристалу, як віднесення до деякого класу структур (наприклад, "об'ємно центрована кубічна решітка", "решітка типу алмазу") тощо. Інформаційними моделями твердого тіла є, наприклад, залежності його фізичних властивостей (питомий електричний опір, період кристалічної решітки, питома теплоємність та інші) від зовнішніх параметрів (температура, тиск, напруженість електричного або магнітного полів тощо), спектри оптичного поглинання, рентгенограми тощо. Інформаційними моделями будуть також взаємозв'язки між різними властивостями, зокрема залежність коефіцієнту дифузії деякої домішки від вмісту інших домішок. Ці моделі можна побудувати у вигляді таблиць даних, графіків або функцій.

Існує багато різних класифікацій математичних моделей. Зокрема, виділяють моделі статичні та динамічні, диференціальні й інтегральні, детерміністичні та стохастичні, лінійні та нелінійні, геометричні, топологічні, імітаційні, оптимізаційні тощо. Найпоширенішими формами запису математичних моделей є інваріантна, аналітична, алгоритмічна та схемна (графічна). В інваріантній формі моделі записують за допомогою алгебраїчних, диференціальних, інтегральних та інших рівнянь і нерівностей без урахування методу подальшого аналізу моделі. Аналітична форма -- це запис моделі у вигляді аналітичного розв'язку вихідних рівнянь інваріантної моделі. Алгоритмічна форма є запис

алгоритму дослідження вихідної моделі, тобто послідовності операцій, що здійснюються при такому дослідженні. Моделі, записані у графічній формі, -- це геометричні та топологічні об'єкти, графи, схеми, графіки тощо.

З погляду практичного використання важливою є класифікація моделей за методом їх подальшого аналізу. При цьому виділяють моделі, які досліджують аналітично, чисельно та за допомогою апаратного моделювання (аналогових обчислювальних машин).

У першому випадку вихідна математична модель має бути перетворена у таку систему співвідношень, яка дає можливість отримати необхідний результат аналітичними методами. Зазвичай результатом аналітичного дослідження є побудова формул, що задають шукані величини в явному вигляді; перетворення рівнянь до вигляду, для якого відомий аналітичний розв'язок, тощо. Результатами аналітичного дослідження також можуть бути якісні висновки про наявність особливих точок, асимптотику, монотонність і однозначність залежності, стійкість розв'язку та інші. Розвиток комп'ютерної техніки та програмного забезпечення дає змогу використовувати для аналітичних досліджень математичних моделей прикладні пакети символьних обчислень, такі як Maple, Mathcad, Reduce тощо, що значно розширює можливості здійснення аналітичних досліджень.

Утім, для отримання аналітичного розв'язку часто необхідно спростити вихідну математичну модель. Типовими прикладами таких спрощень є розкладання складної залежності у ряд Тейлора або Фур'є та врахування лише кількох перших членів цього ряду, нехтування деякими величинами, які вважаються малими, тощо. У більшості реальних випадків математична модель не може бути перетворена до вигляду, який надає можливість отримати аналітичний розв'язок за умови збереження її адекватності. Тому для дослідження моделі використовують інші методи.

Чисельне дослідження має ширшу сферу застосування за аналітичне. Це зумовлено тим, що обчислювальні методи аналізу можуть застосовуватися до ширшого кола математичних моделей. Принциповим недоліком обчислювальних методів є те, що з їх допомогою можна виконувати лише аналіз окремих випадків. Тому розв'язки, що отримують, не будуть повними. Вони характеризують поведінку та властивості систем, що досліджуються, і процесів лише за певних умов, які мають бути чітко визначеними.

Ще однією проблемою при використанні обчислювальних методів може бути некоректність поставленої задачі, тобто відсутність у неї розв'язку, його нестійкість або наявність кількох розв'язків. Умову стійкості можна сформулювати в такому вигляді:

$$\|\delta y\| \rightarrow 0 \quad \text{якщо} \quad \|\delta x\| \rightarrow 0, \quad (8)$$

де δx -- вектор похибок вихідних даних, δy -- вектор похибок результатів.

Нестійкість розв'язку означає, що малі похибки значень вихідних параметрів можуть призводити до великих похибок результатів, що отримуються. У таких випадках застосування обчислювальних методів не має сенсу. Можливі ситуації слабкої стійкості (поганої зумовленості), коли умова (8) формально виконується, але похибка результату є неприпустимо велика.

Прикладами коректних задач можуть бути знаходження коренів полінома; розв'язування системи лінійних алгебраїчних рівнянь у випадку, коли кількість рівнянь дорівнює кількості невідомих, і детермінант матриці, складеної із коефіцієнтів рівнянь, є відмінний від нуля; розв'язування задачі Коші для звичайних диференціальних рівнянь тощо.

Для дослідження математичних моделей, що побудовані у вигляді диференціальних рівнянь та їх систем, застосовують також аналогове моделювання. У цьому разі використовують спеціальні установки, які називають аналоговими обчислювальними

машинами. Як правило, це електричні ланцюги, що містять активні опори, конденсатори, котушки індуктивності та джерела струму. Залежності сили току та напруги від часу у таких ланцюгах описуються різноманітними диференціальними рівняннями. Це дає змогу отримувати розв'язки відповідних рівнянь у чисельному або графічному вигляді, виконуючи вимірювання відповідних електричних величин.

При дослідженні складних систем і процесів використовують також методи, основані на дослідженні їх аналогів. Аналогами називають різні за змістом процеси, що описуються одними й тими самими математичними моделями. Це дає можливість вивчати складні для експериментального дослідження природні, технічні та соціально-економічні системи за допомогою інших методів. Часто існування аналогів зумовлено дією загальніших законів, тому дослідження аналогів сприяє глибшому розумінню систем і процесів, що досліджуються.

Класичними прикладами аналогів є фізичні процеси перенесення -- теплопровідність, дифузія та електропровідність. Густини потоку тепла j_T , потоку частинок j_n та електричного струму j_E підпорядковуються відповідно, першому закону Фур'є для теплопровідності, першому закону Фіка для дифузії та закону Ома:

$$j_T = -\lambda_T \nabla T; \quad j_n = -D_T \nabla n; \quad j_E = -\sigma \nabla \varphi; \quad (9)$$

де λ_T -- коефіцієнт теплопровідності, T -- температура, D -- коефіцієнт дифузії, n -- концентрація дифундуючих частинок, σ -- питома електрична провідність, φ -- потенціал електричного поля.

Для нестационарних процесів також спостерігається аналогія:

$$\frac{\partial T}{\partial t} = a \nabla^2 T; \quad \frac{\partial n}{\partial t} = D \nabla^2 n; \quad \frac{\partial \varphi}{\partial t} = \frac{\sigma}{C^*} \nabla^2 \varphi, \quad (10)$$

де t -- час, a -- коефіцієнт температуропровідності, $C^* = dq/d\varphi$ -- питома електрична ємність матеріалу, q -- електричний заряд.

Закон подібності між теплопровідністю та електричним струмом має вигляд:

$$\frac{\sigma}{C^* a} \frac{t_E}{t_T} \left(\frac{l_T}{l_E} \right)^2 = 1, \quad (11)$$

де l, t -- відповідні довжини та часи. Цей закон містить два коефіцієнти подібності $\frac{t_E}{t_T}$ і

$\frac{l_E}{l_T}$, а також критерій подібності (число Бойкена) $Вей = \frac{\sigma}{C^* a}$.

Закон подібності між теплопровідністю та дифузією має вигляд:

$$\frac{a}{D} \frac{t_T}{t_D} \left(\frac{l_D}{l_T} \right)^2 = 1, \quad (12)$$

У цьому разі коефіцієнтами подібності є $\frac{l_T}{l_D}$ і $\frac{t_D}{t_T}$, а критерієм подібності -- число Льюїса $Le = a/D$.

Розглянемо як приклад математичну модель дифузії домішки у твердому тілі. Вихідну модель у разі вільної дифузії можна записати у такому вигляді:

$$\begin{cases} \frac{\partial n}{\partial t} &= \text{div}(D(r,t)\text{grad}n(r,t); \\ n(r_0,t) &= \varphi(t); \\ n(r,t_0) &= \psi(r), \end{cases} \quad (13)$$

де r -- радіус-вектор.

Модель (7) не має аналітичного розв'язку. Але за деяких умов вона може бути перетворена до моделей, що можна досліджувати аналітично. Наприклад, однією з типових задач дифузії є одновимірна дифузія із необмеженого однорідного плоского джерела у півнескінченне однорідне тіло за сталого коефіцієнту дифузії. Така модель дифузії широко використовується при аналізі багатьох процесів хіміко-термічної обробки металів, процесів створення твердотілих електронних приладів тощо. Для цього випадку, якщо можна нехтувати присутністю домішки в тілі при $t = 0$, модель (7) набирає вигляду:

$$\begin{cases} \frac{\partial n}{\partial t} &= D \frac{d^2 n}{dx^2}; \\ n(0,t) &= n_0; \\ n(x,0) &= n_0 \end{cases} \quad (14)$$

Із системи (14) можна отримати рівняння, що описує концентрацію у довільний момент часу у довільній точці тіла:

$$n(x,t) = n_0 \text{erfc} \frac{x}{2\sqrt{Dt}}, \quad (15)$$

де

$$\text{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx.$$

У дійсності зроблені припущення або не виконуються взагалі, або виконуються лише приблизно. Зокрема, тіло, в яке дифундує домішка, зазвичай буває не зовсім однорідним. Тому коефіцієнт дифузії у точці залежить від її радіус-вектора. Крім того, він залежить також від концентрації домішки і, внаслідок цього, від часу. Реальні тіла завжди скінченні, а джерела домішки обмежені. Якщо цими обставинами нехтувати не можна, модель (13) не має аналітичного розв'язку й її дослідження слід виконувати за допомогою обчислювальних методів.

Загальна схема чисельного дослідження системи (13) може бути такою. На першому етапі задають початковий розподіл концентрації домішки $n(x,0)$, крок за часом Δt і крок за координатою Δx . Це дозволяє обчислити похідну

$$\frac{d^2 n}{dx^2} = \frac{n(x+2\Delta x,0) - 2n(x,0) + n(x-2\Delta x,0)}{4\Delta x^2} \quad (16)$$

для будь-якої точки тіла. За її значеннями можна розрахувати швидкості зміни концентрації у відповідних точках для початкового моменту часу. Після цього знаходять розподіл концентрації домішки для наступного моменту часу:

$$n(x,\Delta t) = n(x,0) + \frac{dn}{dt} \Big|_{t=0} \Delta t. \quad (17)$$

Потім визначають нові значення похідної $\frac{d^2 n}{dx^2} \Big|_{x,t=\Delta t}$, нові швидкості зміни

концентрації в усіх точках і значення концентрації для наступного моменту часу $t = 2\Delta t$. Продовжуючи цю процедуру, можна отримати розподіл домішки у твердому тілі для будь-якого моменту часу.

У загальному вигляді математична модель для розрахунку розподілу температури у твердому тілі може бути такою:

$$\begin{cases} j_T &= -\lambda_T(r,t)\text{grad}T(r,t); \\ \frac{\partial T}{\partial t} &= \text{div}(\lambda_T(r,t)\text{grad}T(r,t)); \\ T(r_0,t) &= \zeta(t); \\ T(r,t_0) &= \zeta(r), \end{cases} \quad (18)$$

Вимірювання розподілу температури у деяких випадках є простіша задача, ніж вимірювання розподілу концентрації домішки. Тоді розподіл концентрації можна знайти таким чином. На першому етапі визначають параметри системи, для якої будуть вимірювати розподіл температури. Потім експериментально визначають розподіл температури. Після цього з використанням критеріїв подібності його перераховують у розподіл концентрації.

Важливий клас математичних моделей складних систем становлять імітаційні моделі. При імітаційному моделюванні відтворюють елементарні явища, що відбуваються в системі, що досліджується, зі збереженням їхньої структури, взаємозв'язків та послідовності протікання. Це дає змогу вивчати розвиток системи у часі, розв'язувати складніші задачі порівняно з аналітичним моделюванням, надає можливість враховувати наявність дискретних і неперервних елементів, нелінійність характеристик елементів системи, випадкові впливи та інші ефекти. Варіантами імітаційного моделювання є метод статистичних випробувань (метод Монте--Карло) і метод статистичного моделювання. Метод Монте--Карло передбачає багаторазове відтворювання процесів, що є реалізації випадкових величин або функцій, та подальшу статистичну обробку отриманих результатів. Він використовується для моделювання випадкових величин та функцій, статистичні характеристики яких попередньо були отримані як розв'язки аналітичних задач. Метод статистичного моделювання використовується для дослідження випадкових впливів на характеристики систем або процесів, що досліджуються.

1.1.3 Основні властивості математичних моделей

До основних властивостей математичних моделей належать їх скінченність, спрощеність, наближеність, повнота, адекватність та істинність.

Скінченність моделі означає, що вона відображає лише деякі з характеристик та відношень, властивих оригіналу. Її зумовлено обмеженістю ресурсів часу, пам'яті ЕОМ тощо, потрібних для розробки й аналізу моделі.

Спрощеність моделі означає, що у деякому сенсі вона має бути простішою за оригінал. У цьому полягає сенс моделювання, оскільки у противному разі доцільніше досліджувати оригінал. Спрощення досягають шляхом нехтування другорядними, несуттєвими для досягнення мети моделювання властивостями оригіналу. При побудові моделей використовують принцип руху від простого до складного. Тобто, спочатку будують найпростішу можливу модель системи (як правило, такі моделі можна знайти у фаховій літературі). Потім перевіряють її адекватність. Якщо простіша модель неадекватна, то визначають, чим зумовлено її неадекватність. Зазвичай причинами є

нехтування у змістовній моделі деякими факторами, що істотно впливають на досліджувану систему, або надмірна спрощеність математичної моделі.

Типовими прикладами наближень, які використовують при математичному моделюванні, є заміна дискретних систем неперервними та навпаки, завдання обмежень на точність чисельних розрахунків, заміна нелінійних залежностей лінійними тощо. Ступінь наближеності моделі визначається компромісом між необхідністю відображення всіх суттєвих властивостей оригіналу й обмеженістю часу, пам'яті ЕОМ та інших ресурсів.

Зі скінченності та наближеності моделі випливає, що вона відображає оригінал неповно. Ступінь повноти моделі визначається метою та завданнями моделювання.

Адекватність моделі характеризує можливість реалізації мети моделювання, а істинність -- її відповідність сукупності наявних знань про об'єкт дослідження. Критеріями адекватності є відображення всіх істотних властивостей та параметрів об'єкту дослідження, якісно правильне відображення суттєвих зв'язків між параметрами, а при кількісному дослідженні також мала різниця між результатами моделювання та наявними емпіричними даними. Адекватність є обов'язкова вимога до будь-якої моделі, що використовується для дослідження реальних систем. Але у деяких випадках застосовують відносно прості моделі, що не є адекватні. Це доцільно, зокрема, під час розробки алгоритмів аналізу моделей (на першому етапі розробляють алгоритми аналізу неадекватних моделей, які потім використовують як елементи більш складної моделі), у навчальному процесі (зокрема, коли треба дослідити явища, які у реальних системах та процесах спотворюються іншими ефектами) тощо. Слід пам'ятати, що істинність моделі не є гарантія її адекватності. Зокрема, це може бути зумовлено накопиченням похибок за необхідності виконання великого обсягу розрахунків. З іншого боку, адекватними можуть бути моделі, що не є істинні. Типовими прикладами є регресійні моделі, які дають змогу прогнозувати поведінку досліджуваної системи у деякому діапазоні зміни вхідних параметрів, але не відображають наявних знань про побудову, зв'язки та внутрішні процеси у ній.

1.1.4 Загальна схема математичного моделювання

Загальна схема побудови математичної моделі є така. Насамперед необхідно визначити, для чого необхідна модель. Це має принципове значення для обрання методу її побудови. Існують два основні підходи. У першому випадку математична модель ґрунтується на відомих теоретичних даних про закономірності поведінки системи або протікання процесу. У цьому разі математична модель, що отримується, буде системою відомих з теорії моделей. Перевагами такого підходу є відповідність структури моделі реальній структурі об'єкта дослідження. Завдяки цьому всі параметри моделі мають реальний фізичний, економічний, технічний або інший зміст. Такі моделі дають змогу аналізувати не тільки загальні властивості системи як цілого, але також і поведінку окремих її елементів, зміни структури, визначати відносні внески різних факторів у властивості, що спостерігаються, тощо. Недоліками цього підходу зазвичай є складність цих моделей і, внаслідок цього, можливість накопичення похибок при розрахунку вихідних характеристик об'єкта дослідження.

У другому випадку моделлю є рівняння регресії (або система таких рівнянь), за допомогою якого можна прогнозувати, як будуть змінюватися характеристики системи або процесу при зміні вхідних змінних. Рівняння регресії відповідає розгляду системи як чорної скриньки. Тому такі моделі принципово неможливо використовувати для оцінювання внеску окремих підсистем у формування загальних властивостей системи, аналізу структури системи, зв'язків між її елементами тощо. Коефіцієнти рівнянь регресії

часто не мають якогось реального змісту. Проте регресійні моделі відрізняються простотою й у багатьох випадках дають можливість отримувати точніші оцінки вихідних характеристик систем та процесів, що досліджуються.

У багатьох випадках при побудові моделей складних систем і процесів використовують комбінації цих підходів. Зокрема, часто базову модель будують як систему відомих теоретичних моделей, параметри яких визначають за допомогою регресійного аналізу.

Якщо модель будують у вигляді системи відомих теоретичних моделей, наступним етапом її розробки є визначення концептуальної або змістовної (фізичної, технічної, економічної тощо) моделі об'єкта дослідження. Змістовну модель будують на основі відомих теоретичних та емпіричних даних про об'єкт. На цьому етапі визначають суттєві для задачі, що розв'язується, елементи системи, взаємозв'язки між ними, взаємозв'язки системи й навколишнього середовища, можливі стани системи, закономірності поведінки системи у цілому та її окремих елементів тощо. Потім переходять від змістовного до формального опису, тобто відбирають теоретичні моделі, із яких будуватиметься загальна математична модель об'єкта дослідження, визначають межі застосування зроблених у них припущень і спрощень.

Далі процедура розробки моделі залежить від обрання методики її подальшого аналізу. На сьогодні найбільш поширеним методом дослідження математичних моделей є їх чисельний аналіз за допомогою ПК. Для цього можна використовувати різноманітні математичні, статистичні та інші прикладні пакети програмного забезпечення, зокрема Microsoft Excel, MathCad, MathLab, Mathematica, Statistica тощо. Але слід мати на увазі, що не існує алгоритмів чисельних розрахунків, які б давали змогу отримати задовільні розв'язки для всіх задач деякого класу. Якість роботи алгоритму залежить не тільки від типу задачі, але й від її конкретних умов та параметрів. Тому обрання алгоритму є нетривіальна задача. Як правило, якість алгоритму (й моделі взагалі) визначають порівнянням результатів, що отримують за різними алгоритмами розрахунків, зокрема виконанням розрахунків для моделей із відомими характеристиками, а також порівнянням результатів моделювання із відомими даними про систему, що моделюють (якісними закономірностями її поведінки, кількісними значеннями її параметрів тощо). У типових прикладних пакетах, як правило, не вказується конкретний алгоритм, за яким виконуються розрахунки. Більше того, досить часто не вказується також математичний метод, на якому базується цей алгоритм. Це суттєво ускладнює попередній аналіз можливості застосування прикладних пакетів у дослідженні тієї чи іншої моделі, а також пошук джерел похибок моделювання.

Важливою характеристикою результатів моделювання є похибка результатів, що отримують. Можна виділити такі її основні складові: похибка математичної моделі, похибка вихідних даних, похибка розрахункового алгоритму та похибка обчислень.

Похибка математичної моделі пов'язана із тим, що будь-яка модель є лише наближене відображення об'єкта моделювання. Джерелами цієї похибки є припущення, що роблять при розробці змістовної моделі об'єкта дослідження, а також наближеність при побудові її математичної моделі. Зокрема, наближеність моделі ((13)) визначається припущеннями про однорідність твердого тіла, незалежність коефіцієнту дифузії від концентрації домішки та від часу, нехтуванням впливу внутрішніх та зовнішніх електромагнітних полів на поведінку електрично заряджених домішок тощо. Якщо необхідно, похибку моделі можна зменшити шляхом урахування додаткових факторів, що впливають на вимірювані характеристики.

Похибка вихідних даних пов'язана із похибками вимірювань, використанням наближених значень параметрів системи, заміною генеральних сукупностей вибірками

обмеженого обсягу тощо. Вона може бути зменшена тільки шляхом виконання повторних або додаткових експериментів та спостережень.

Похибка розрахункового алгоритму пов'язана зі спрощеннями та припущеннями, що роблять при заміні вихідної математичної моделі алгоритмом обчислень. Зокрема, більшість методів чисельного знаходження інтегралів використовує заміну площі під кривою, що інтегрується, на суму площ відносно простих фігур. При цьому похибка інтегрування суттєво залежить від того, які саме фігури буде обрано для наближення, а також від кількості ділянок, на які розбивають інтервал інтегрування. Багато алгоритмів використовують ітераційні процедури, що теоретично можуть виконуватися нескінченно довго, поступово наближуючись до точного розв'язку. Але у практиці час виконання ітерацій має бути обмеженим. Тому необхідно заздалегідь встановити точність результату, яка задовольняє дослідника.

Похибка обчислень пов'язана з необхідністю обмеження кількості значущих цифр у числах, з якими виконують розрахунки.

Похибки розрахункового алгоритму та обчислень належать до усувних похибок, оскільки вони можуть бути зменшені до необхідного рівня шляхом внесення змін до алгоритму та підвищення точності розрахунків.

Похибки вихідних даних та моделі належать до неусувних похибок. Це треба розуміти так: ці похибки не можна усунути або зменшити математичними методами, але точність моделювання може бути підвищена шляхом обрання адекватнішої моделі (цей етап є неформальний, і для його реалізації необхідно використовувати не тільки математичні методи, а й дані відповідних конкретних наук), а також підвищення точності вихідних даних.

При побудові моделі необхідно прагнути до балансу похибок різного типу. Це зумовлено тим, що похибка моделювання визначається найбільшою із них. Підвищувати точність розрахунків на деякому етапі моделювання недоцільно, якщо вихідні дані для нього мають суттєво меншу точність або якщо результати цього етапу потім використовують для розрахунків зі значно більшою похибкою. Можна рекомендувати таку послідовність балансу похибок. На першому етапі необхідно визначити похибку вихідних емпіричних даних, що використовуватимуться при розрахунках, та характер впливу цієї похибки на результати моделювання. Загалом похибка результатів перевищує похибку вихідних даних і є тим вища, що складніша модель. Але при цьому слід розрізняти систематичну та випадкову похибку моделювання. У деяких випадках похибка визначення сталих коефіцієнтів, що входять до моделі, призводить лише до появи систематичної похибки. Якщо цей коефіцієнт використовують у моделі як самостійний доданок, то його похибка зумовлює збільшення або зменшення результату на одну й ту саму величину. При цьому якісна поведінка моделі та всі інші її кількісні характеристики не змінюються. Якщо коефіцієнт використовують як загальний множник, то відхилення вихідних результатів будуть пропорційні похибці його визначення. При цьому якісна поведінка моделі також залишається правильною. У складніших випадках похибки вихідних даних можуть призводити до якісних змін у поведінці моделі.

Розглянемо, наприклад, модель, задану у вигляді диференційного рівняння

$$\frac{dy}{dx} = ax^2 + bx + c, \quad (19)$$

для якої емпірично встановлено такі значення коефіцієнтів: $a = 2; b = 4; c = 2$. Для цих значень величина $dy/dx = 0$ у точці $x = -1$, в інших точках вона більша за нуль. Відповідно, функція $y(x)$ є зростаюча і має одну точку перегину, в якій її похідна дорівнює нулю. Істинні значення коефіцієнтів моделі відрізняються від емпірично

встановлених. Нехай, наприклад, вони становлять $a = 1,9; b = 4,1; c = 1,9$. Тоді маємо дві точки $x_1 \approx -1,484$ та $x_2 \approx -0,674$, де похідна дорівнює нулю. При $x \in (-\infty, x_1) \cup (x_2, \infty)$ функція зростає, а при $x \in [x_1, x_2]$ -- спадає. Відповідно, x_1 є точка максимуму, а x_2 -- точка мінімуму функції $y(x)$. Якщо істинні значення коефіцієнтів моделі становлять $a = 2,14; b = 3,9; c = 1,9$, похідна є додатна для всіх x , і функція не має точок екстремуму. Така різниця поведінки функції $y(x)$ може бути істотною з точки зору змістовної моделі системи. Зокрема, якщо $y(x)$ -- потенційна енергія фізичної системи, то ми отримуємо три суттєво різні фізичні моделі. Залежно від істинних значень коефіцієнтів система може мати точки стійкої та нестійкої рівноваги, тільки точку нестійкої рівноваги або ж взагалі не мати точок рівноваги. Якщо похибка вихідних даних може призводити до якісно різних результатів моделювання, подальший аналіз такої моделі має сенс лише у випадках, коли ставиться завдання визначити можливі сценарії розвитку системи, її можливі стани тощо. Інакше необхідно підвищити точність вихідних даних, побудувати іншу модель або змінити завдання дослідження. Якщо похибка вихідних даних дає можливість досягти мети моделювання, наступним етапом є оцінка та порівняння похибок проміжних і підсумкових розрахунків. За результатами порівняння робиться висновок про збалансованість похибок, необхідність доробки алгоритму чи підвищення точності розрахунків на деякому етапі, якщо точність результатів не задовольняє дослідника, або про можливість спрощення алгоритму чи зменшення точності деяких розрахунків, коли точність, що отримується на відповідному етапі, є надмірна.

Після побудови математичної моделі треба визначити її адекватність. Для цього треба порівняти результати моделювання із емпіричними даними, що отримано за близьких умов, а також із результатами, отриманими на інших моделях. Інколи для перевірки адекватності треба залучати незалежних експертів, які не брали участі у розробці моделі.

За результатами перевірки адекватності моделі приймають рішення щодо можливості її використання. Результатом перевірки може бути рішення про необхідність доробки (корегування) та оптимізації моделі. При корегуванні уточнюють перелік суттєвих параметрів моделі, обмеження, функціональні зв'язки між параметрами тощо. Під оптимізацією розуміють спрощення моделі при збереженні заданого рівня її адекватності. Основними критеріями оптимальності є витрати часу, пам'яті ЕОМ та інших ресурсів при використанні моделі.

1.1.5 Поняття моделі та способи побудови моделей

Модель -- спрощене подання об'єкту, системи чи поняття у деякій абстрактній формі, що є зручна для наукового дослідження.

У загальному випадку модель має структуру, зображену на рис. 1. Тут X -- множина вхідних змінних системи, Y -- множина вихідних змінних системи, P --- множина параметрів, F -- функція, функціонал, алгоритм або формальне представлення залежності змінних Y від змінних X .

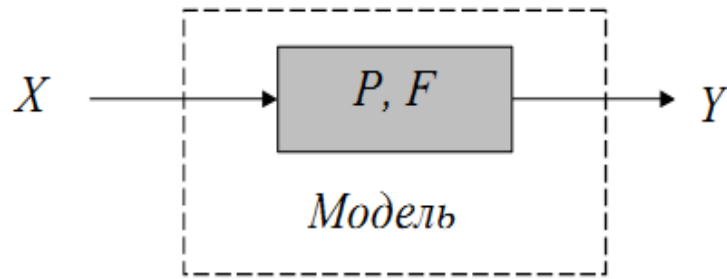


Рис. 1 Загальна структура моделі

Наприклад, моделі масового обслуговування описуються набором вхідних змінних, що складається із інтенсивності вхідного потоку на обслуговування, тривалості обслуговування кожним пристроєм, ймовірностей вибору маршруту. Набір вихідних змінних моделі масового обслуговування складається з таких змінних як інтенсивність вихідного потоку, середня довжина черги у місцях накопичення та інших. В якості параметрів моделі масового обслуговування можуть розглядатись кількість пристроїв обслуговування, наявність черг у місцях накопичення вимог, наявність обмежень на кількість місць у чергах та інші. В якості алгоритму -- алгоритм розрахунку вихідних змінних моделі за значенням вхідних змінних, який залежно від складності моделі може мати вид простого розрахунку за математичними формулами або вид алгоритму імітації системи.

Способи побудови моделей

Існують два способи побудови моделей. При першому способі у результаті ретельного вивчення системи встановлюються закони функціонування системи, які потім відтворюються за допомогою моделі. Поведінку системи, таким чином, досліджують на моделі. Параметри P моделі у цьому випадку пов'язані з реальними процесами, що протікають у системі, і мають фізичну інтерпретацію. Тому моделі такого типу називають фізичними моделями.

При другому без усякого фізичного обґрунтування припускається вид залежності F , невідомі параметри якої P потім відшукуються за даними спостережень за змінними системи X, Y . Параметри P у цьому разі не зв'язані з фізикою реальних процесів, що протікають у системі, або, точніше, цей зв'язок досліднику залишається невідомим. Тому моделі такого типу називають нефізичними моделями.

В літературі трапляються також терміни моделей типу сірої та чорної скриньки, які еквівалентні термінам фізичної та нефізичної моделі. Для фізичної моделі закони функціонування системи досліднику відомі, тому скринька є прозора, сіра. Для нефізичної моделі сутність системи залишається для дослідника прихованою, потаємною, тобто скринька є чорна.

Задачі моделювання

Серед задач моделювання виділимо такі задачі: задача моделювання, задача управління, задача ідентифікації, задача оптимізації, задача прогнозування.

Задача *моделювання* (або пряма задача) полягає у відшуванні значень вихідних змінних Y при відомих значеннях вхідних змінних X , відомій моделі F та визначених

параметрах P (див. рис 2).

Задача *управління* (або зворотна задача) полягає у відшукуванні таких значень вхідних змінних X , що забезпечують задані значення вихідних змінних Y при відомій моделі F та фіксованих значеннях параметрів P .

У формулюванні задачі *ідентифікації* відомими являються множина вхідних змінних X , множина вихідних змінних Y та множина моделей F . Потрібно визначити єдину модель f із множини запропонованих моделей F , і визначити її параметри P , що забезпечують при вхідних значеннях X вихідні значення Y .

У постановці задачі *оптимізації* відомими являються модель F , множина можливих вхідних значень X та критерій оптимізації K , а від дослідника вимагається знайти значення вхідних змінних X , значення параметрів P та значення вихідних змінних Y , що задовольняють заданому критерію оптимізації K .

Задача *прогнозування* формулюється так, що при відомих для дослідника значеннях вхідних та вихідних значень моделі X_t, Y_t до моменту часу t та заданому часі прогнозування T потрібно визначити модель F та її параметри P , які забезпечують найліпший прогноз Y_{t+T} .

Стисло визначення методів моделювання представлені на рис. 2.

З А Д А Ч І М О Д Е Л Ю В А Н Н Я

Моделювання:	відомі $X, P, F \Rightarrow$ знайти Y
Управління:	відомі $Y, P, F \Rightarrow$ знайти X
Ідентифікації:	відомі X, Y , множина $F \Rightarrow$ знайти $f \in F, P$
Оптимізації:	відомі F , критерій $K \Rightarrow$ знайти P, X, Y
Прогнозування:	відомі $X_t, Y_t, T \Rightarrow$ знайти F, P, Y_{t+T}

Рис. 2 Задачі моделювання

1.1.6 Методи і процес моделювання

Серед великої кількості методів моделювання, що існують, виділимо такі методи: аналітичне моделювання, математичне моделювання, імітаційне моделювання.

Моделювання належить до аналітичного, якщо представлення залежності F вихідних змінних Y від вхідних її змінних X має аналітичний вигляд, тобто представлений у вигляді відомих аналітичних функцій. Нагадаємо, що функція називається аналітичною, якщо вона розкладається у ряд Тейлора. Аналітичні функції диференційовні безліч разів і тому до них можуть застосовуватись методи математичного аналізу. Перевагою цього методу моделювання є можливість отримання залежності $y = f(x)$ в явному вигляді і застосування до неї методів класичного математичного аналізу. Якщо є можливість побудувати аналітичну модель системи, то завжди віддають

перевагу цьому методу моделювання. Зауважимо, що відшукування залежності $y = f(x)$ може виявитись настільки складним, що досліднику доведеться застосовувати спеціальне програмне забезпечення, а для деяких систем доводиться відмовлятися від пошуку абстрактної залежності $y = f(x)$ і задовольнятися наближеним розв'язком, що відшукують чисельними методами.

Деякі системи настільки складні, що не дивлячись на те, що опис їх функціонування піддається опису аналітичними функціями, знаходження залежності $y = f(x)$ в явному вигляді виявляється неможливим. Наприклад, усі задачі математичного програмування мають досить простий аналітичний опис, але розв'язок задачі може бути знайдений тільки у результаті виконання певної кількості кроків. Іншими словами, відомий алгоритм відшукування точного розв'язку задачі, але сам розв'язок не може бути записаний в аналітичній формі. Такий метод моделювання називають *математичним моделюванням*. Зауважимо, що алгоритм F відшукування точного розв'язку задачі може бути реалізований дослідником самостійно, за допомогою спеціального програмного забезпечення або за допомогою чисельних методів.

Існують системи, опис яких не піддається опису аналітичними функціями, але процес функціонування їх може бути описаний алгоритмом імітації. Під імітацією розуміють відтворення за допомогою комп'ютерної програми процесу функціонування складної системи у часі. У результаті багатократних прогонів імітаційної моделі дослідник отримує інформацію про властивості реальної системи. Такий метод моделювання називають імітаційним моделюванням.

Стисло визначення методів моделювання представлені на рис.3.

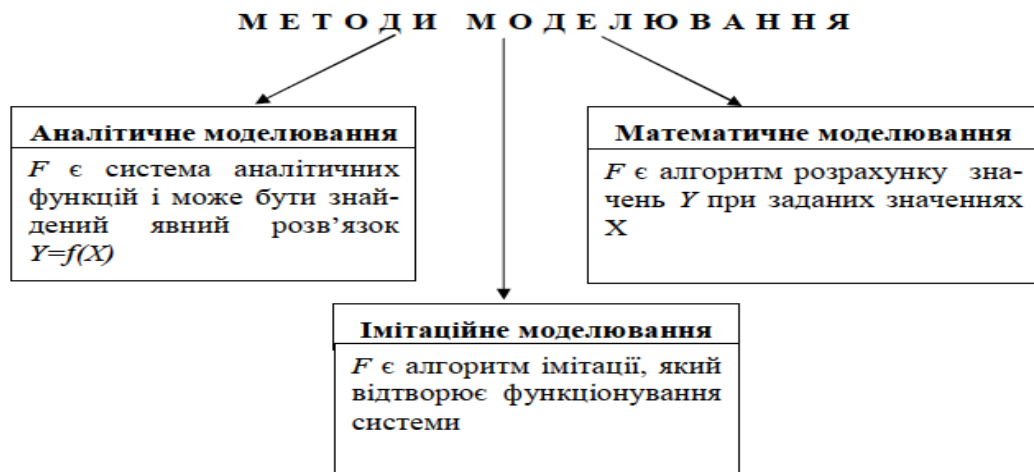


Рис.3 Методи моделювання

Процес моделювання

Процес моделювання складається з кількох етапів. На першому етапі дослідник визначає мету та задачу моделювання. На другому етапі, виходячи з мети та задачі моделювання, дослідник приступає до вербального опису системи. Опис набору змінних моделі, разом із описом структури системи та формулюванням цілі та задачі дослідження складає концептуальну модель системи. Виходячи з концептуальної моделі системи та з огляду на вибір інструментальних засобів, дослідник робить вибір теоретичної бази, на основі якої буде побудована модель системи. Отже, обравши теоретичну базу моделювання, дослідник має описати систему, що розглядається, обраними елементами

формального опису і визначити для них усі необхідні параметри. Формальне представлення системи має вигляд схеми, в якій указані зв'язки між елементами системи та зв'язки із зовнішнім середовищем й указані параметри елементів системи. У формальній моделі міститься також інформація, яким чином будуть знайдені вихідні змінні моделі в результаті моделювання. Наприклад, якщо в якості теоретичної бази моделювання обрані засоби мереж масового обслуговування, формальна модель представляється зображенням мережі масового обслуговування, що складена за умовою задачі, з визначенням числових значень вхідних змінних і параметрів, а також формули розрахунку вихідних змінних моделі, що є мета моделювання.

На третьому етапі дослідник приступає до створення моделі. Спочатку виконується реалізація моделі за допомогою обраного програмного забезпечення. Потім виконується верифікація моделі, тобто перевірка алгоритму моделювання на відповідність задуму моделювання. Наприклад, змінюють значення вхідних змінних і спостерігають як модель реагує на таке змінювання. Якщо реакція моделі відповідає логіці її функціонування, то модель вважається правильною. Завершується створення моделі перевіркою адекватності моделі, що полягає у порівнянні значень вихідних змінних об'єкта, що моделюється, і моделі при однакових значеннях вхідних змінних. Очевидно, що таку перевірку можна здійснити тільки якщо відомі деякі значення вхідних і вихідних змінних об'єкта.

Четвертий етап -- це дослідження моделі. Результати моделювання стають корисними, якщо проведене змістовне дослідження моделі відповідно до мети моделювання. Експерименти, що проводяться з моделлю, мають бути спочатку сплановані, потім -- проведені, і наприкінці -- статистично оброблені. Наприклад, якщо при дослідженні технологічного процесу обробки деталей була поставлена мета -- виявлення місць накопичення деталей, то у результаті моделювання слід не тільки указати ці місця накопичення та обсяги накопичення, але й дослідити, які фактори впливають на зменшення накопичення деталей, і запропонувати заходи щодо зменшення обсягів накопичення деталей.

Аналіз результатів моделювання складається із оцінки точності результатів моделювання, оцінки стійкості результатів моделювання та оцінки чутливості результатів моделювання.

Формування висновків та пропозицій є завершальний етап моделювання, на якому підводяться підсумки та висловлюються думки щодо напрямків подальшого дослідження об'єкту моделювання.

Звісно, що процес моделювання може бути представлений етапами тільки у звіті про результати моделювання. У ході моделювання досліднику доводиться неодноразово повертатись до попередніх етапів і уточнювати постановку задачі, формальний опис моделі, алгоритм реалізації або план проведення експериментів, поступового наближаючись до мети.

ПРОЦЕС МОДЕЛЮВАННЯ

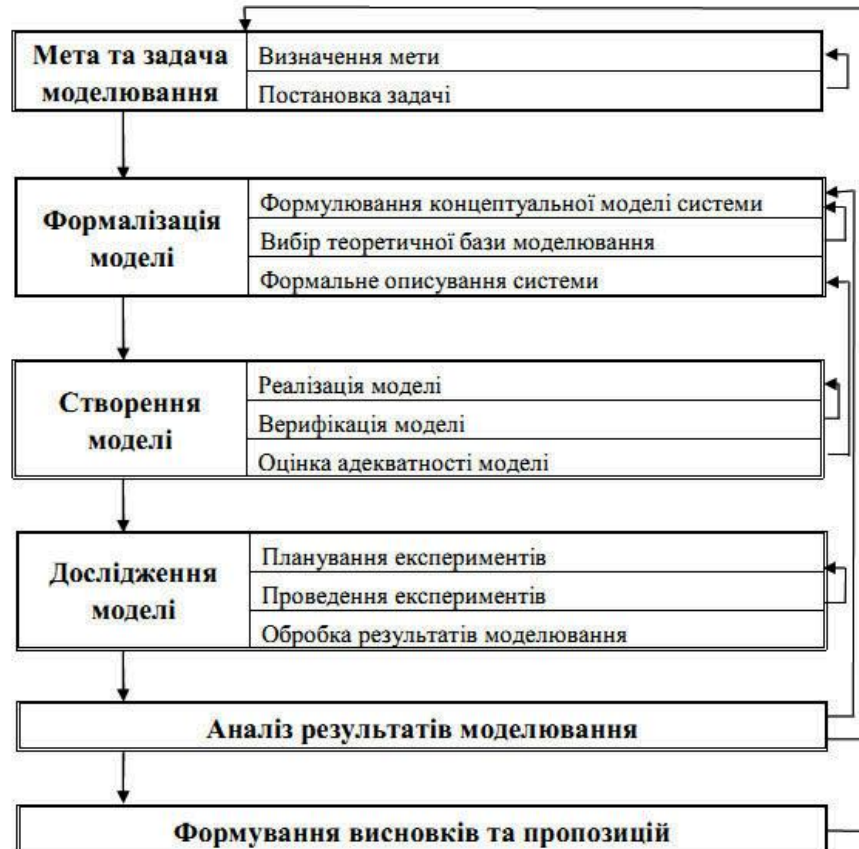


Рис.4 Процес моделювання

1.1.7 Приклади розв'язання задач

Задача 1. Запропонуйте множину вхідних змінних та множину вихідних змінних моделі, а також множину параметрів моделі, якщо об'єкт -- технологічний процес обробки деталей.

Розв'язання. Функціонування технологічного процесу обробки деталей залежить, насамперед, від інтенсивності потоку деталей, що надходять на обробку. Структура технологічного процесу описується кількістю та послідовністю технологічних операцій, характеристиками обладнання, призначеного для виконання кожної операції технологічного процесу, кількістю працівників, зайнятих у кожній операції. До характеристик обладнання відносять такі величини, як тривалість виконання технологічної операції, кількість деталей, що одночасно оброблюються, кількість пристроїв, що призначені для виконання технологічної операції. Складне обладнання, як правило, часто псується у процесі експлуатації, тому слід передбачити у моделі можливість виходу з ладу обладнання й описати процес виникнення та усунення несправностей обладнання.

Основним показником ефективності технологічного процесу є інтенсивність вихідного потоку деталей. Важливими показниками є також коефіцієнт завантаження обладнання, завантаження працівників, зайнятих у технологічному процесі. У деяких задачах розглядається проблема накопичення деталей, що на даний момент не можуть опрацюватись технологічним процесом.

Відповідь: Множина вхідних змінних технологічного процесу обробки деталей може складатись залежно від цілі моделювання з таких змінних:

- інтенсивність надходження деталей на початок технологічного процесу,
- кількість операцій технологічного процесу,
- кількість та характеристики обладнання, призначеного для виконання кожної операції технологічного процесу,
- кількість працівників, зайнятих у кожній операції технологічного процесу;
- ймовірності виходу з ладу обладнання,
- тривалість виконання кожної операції технологічного процесу,
- кількість місць для зберігання деталей, що на даний момент не можуть опрацьовуватись технологічним процесом -- для кожної операції технологічного процесу окремо.

Множина вихідних змінних технологічного процесу обробки деталей може складатись залежно від цілі моделювання із таких змінних:

- інтенсивність вихідного потоку деталей на кінець технологічного процесу,
- завантаження обладнання,
- завантаження працівників,
- обсяги накопичення деталей у місцях зберігання.

В якості параметрів технологічного процесу обробки деталей залежно від цілі моделювання можуть розглядатись:

- кількість операцій технологічного процесу;
- кількість обладнання, що забезпечує виконання технологічного процесу;
- кількість працівників, зайнятих у технологічному процесі;
- наявність місць для зберігання деталей, що на даний момент не можуть опрацьовуватись технологічним процесом.

Задача 2. Запропонуйте множину вхідних змінних та множину вихідних змінних моделі, а також множину параметрів моделі, якщо об'єкт -- система управління транспортним рухом на перехресті.

Відповідь: Множина вхідних змінних системи управління транспортним рухом на перехресті може складатись залежно від цілі моделювання із таких змінних:

- інтенсивність надходження автомобілів на ділянку транспортного руху,
- ймовірності вибору маршруту слідування автомобіля.

Множина вихідних змінних системи управління транспортним рухом на перехресті може складатись залежно від цілі моделювання із таких змінних:

- обсяги накопичення автомобілів на кожному напрямку руху,
- середній час очікування автомобіля на кожному напрямку руху,
- максимальний час очікування автомобіля на перехресті.

В якості параметрів системи управління транспортним рухом залежно від мети моделювання можуть розглядатися:

- структура перехрестя (Т-образне перехрестя, Х-образне, кільце),
- кількість напрямків руху,
- кількість рядів руху для кожного напрямку руху,
- параметри управління транспортним рухом -- наявність керування світлофорними об'єктами, наявність дорожніх знаків,
- кількість світлофорних об'єктів та їх характеристики -- кількість фаз управління та їх структура, тривалість горіння червоного, жовтого та зеленого світла у кожному напрямку впродовж кожної фази,
- параметри автомобілів, що надходять, -- їх габарити, маршрути слідування, швидкість руху.

Задача 3. Поставте задачу моделювання для системи обслуговування клієнтів у банку з метою визначення кількості клієнтів, що перебувають у приміщенні банку.

Відповідь: Для відомих значень інтенсивності надходження клієнтів у банк, інтенсивності обслуговування клієнтів у касирів банку та у клерків залежно від виду банківської операції, ймовірності звернення клієнта до касира або до клерка знайти середню та найбільшу кількість клієнтів у банку, що обслуговуються або очікують обслуговування, якщо відомі кількість касирів, кількість клерків, кількість різних банківських операцій та ймовірності їх затребування клієнтом банку, а також обслуговування клієнтів у банку.

1.2 Контрольні запитання

1. Дати визначення моделі системи.
2. Перелічити види моделей системи.
3. Як класифіуються моделі?
3. Назвіть способи побудови моделей систем.
4. Які відомі методи моделювання?
5. У чому полягають переваги імітаційного моделювання систем, які недоліки?
6. Сформулюйте постановку задачі моделювання: задачі оптимізації; задачі ідентифікації; задачі управління; задачі прогнозування.
7. Сформулюйте сутність системного підходу до математичного моделювання.
7. Що таке "концептуальна" модель системи?
8. Перелічити етапи процесу моделювання.
9. Що означає термін "адекватність" моделі?
10. У чому сутність методу математичного моделювання?

Розділ 2. Статистичне моделювання систем

2.1 Вступ

Поняття "статистичне моделювання систем" охоплює досить широке коло задач моделювання. У цьому розділі ми будемо розглядати питання моделювання процесів і сигналів, яким властиві ймовірнісні характеристики із визначеними або ні ймовірнісними законами розподілу.

Сутність методу статистичного моделювання полягає у побудові певного алгоритму для процесу функціонування системи, що досліджується, що імітує поведінку і взаємозв'язок елементів системи із урахуванням випадкових вхідних впливів та впливів зовнішнього середовища, та реалізації цього алгоритму засобами обчислювальної техніки.

Теоретичним підґрунтям методу статистичного моделювання систем є граничні теореми теорії ймовірностей. Множини випадкових явищ (подій, величин тощо) підпорядковуються визначеним закономірностям, що надають можливість як прогнозувати їхню поведінку, так і кількісно оцінити їхні певні характеристики. Принципове значення граничних теорем полягає у тому, що вони гарантують високу якість статистичних оцінок за досить великої кількості дослідів (реалізацій). Практично достатні при статистичному моделюванні кількісні оцінки можна отримати за порівняно невеликих реалізацій.

2.1.1 Граничні теореми теорії ймовірностей

Нерівність Чебишова. Для невід'ємної функції $g(\xi)$ випадкової величини ξ і будь-якого K виконується нерівність

$$P\{g(\xi) \geq K\} \leq M[g(\xi)]/K, \quad (1)$$

зокрема, коли $g(\xi) = (\xi - \bar{x})^2$, $K = k^2 \sigma^2$ ($\bar{x}$ -- середнє арифметичне, σ -- середнє квадратичне відхилення,

$$P\{|\xi - \bar{x}| \geq k\sigma\} \leq 1/k^2. \quad (2)$$

Теорема Бернуллі. Якщо проводиться N незалежних випробувань, у кожному з яких деяка подія A здійснюється із ймовірністю p , відносна частота появи події m/N (m -- кількість додатних витоків випробувань), за $N \rightarrow \infty$ збігається за ймовірністю до p , тобто за будь-якого $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\{|m/N - p| \geq \varepsilon\} = 0. \quad (3)$$

Теорема Пуассона. Якщо проводиться N незалежних випробувань і ймовірність появи події A в i -ому випробуванні дорівнює p_i , відносна частота m/N появи події p за $N \rightarrow \infty$ збігається за ймовірністю до середнього із ймовірностей p_i , тобто

$$\lim_{N \rightarrow \infty} P\{|\frac{m}{N} - \frac{1}{N} \sum_{i=1}^N p_i| \geq \varepsilon\} = 0. \quad (4)$$

Узагальнена теорема Чебишова. Якщо $\xi_1, \dots, \xi_N$ -- незалежні випадкові величини із математичними сподіваннями $a_1, \dots, a_N$ та дисперсіями $\sigma_1^2, \dots, \sigma_N^2$, обмеженими зверху одним і тим самим числом, за $N \rightarrow \infty$ середнє арифметичне значень випадкової величини збігається за ймовірністю до середнього арифметичного їхніх математичних сподівань:

$$\lim_{N \rightarrow \infty} P\left\{\left|\frac{1}{N} \sum_{i=1}^N x_i - \frac{1}{N} \sum_{i=1}^N a_i\right| \geq \varepsilon\right\} = 0. \quad (5)$$

Теорема Маркова. Вираз (5) справджується і для залежних випадкових величин $\xi_1, \dots, \xi_N$, якщо

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} D\left[\sum_{i=1}^N x_i\right] = 0.$$

Сукупність теорем, що встановлюють стійкість середніх показників, прийнято називати *законом великих чисел*.

Центральна гранична теорема. Якщо $\xi_1, \dots, \xi_N$ -- незалежні однаково розподілені випадкові величини із математичним сподіванням a і дисперсією σ^2 , за $N \rightarrow \infty$ закон розподілу суми $\sum_{i=1}^N x_i$ необмежено наближається до нормального:

$$\lim_{N \rightarrow \infty} P\left\{x < \frac{\sum_{i=1}^N x_i - Na}{\sqrt{Na}} < \beta\right\} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2} dt = \Phi_0(\beta) - \Phi_0(\alpha). \quad (6)$$

$$\Phi_0(\gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{-t^2/2} dt.$$

Теорема Лапласа. Якщо у кожному із N незалежних випробувань подія A з'являється із ймовірністю p ,

$$\lim_{N \rightarrow \infty} P\left\{\alpha < \frac{m - Np}{\sqrt{Np(1-p)}} < \beta\right\} = \Phi_0(\beta) - \Phi_0(\alpha). \quad (7)$$

де m -- число появ події A в N випробуваннях. Теорема Лапласа є частинний випадок центральної граничної теореми.

Статистичне моделювання систем на ЕОМ потребує формування значень випадкових величин за допомогою генераторів випадкових чисел. Залежно від характеру (природи) системи, що моделюється, генератори випадкових (псевдовипадкових) чисел можуть бути реалізовані із урахуванням ймовірнісних законів розподілу випадкових величин і процесів.

Кількість випадкових чисел, що використовують для отримання статистично стійкої оцінки характеристики процесу функціонування системи при реалізації алгоритму моделювання, коливається у досить широких межах залежно від класу об'єкту моделювання, виду характеристики, що оцінюються, потрібної точності і достовірності результатів. На ЕОМ неможливо отримати ідеальну послідовність випадкових чисел, оскільки є можливість оперувати тільки зі скінченною множиною чисел. Крім того, для отримання значень x випадкової величини ξ використовуються формули (алгоритми). Тому такі послідовності, що є за сутністю детермінованими, називаються *псевдовипадковими*.

2.2 Комп'ютерне моделювання ПВЧ

Комп'ютерне моделювання псевдовипадкових чисел (ПВЧ) безпосередньо пов'язане із моделюванням випадкових процесів і сигналів довільної природи. ПВЧ використовують для імітаційного моделювання з метою наближення процесів, що

моделюються засобами обчислювальної техніки, до реальних випадкових процесів.

Більшість теоретичних праць, що пов'язані із описом ймовірнісних процесів, розглядають обмежене коло законів розподілу ймовірностей цих процесів, обмежуючись найпопулярнішим нормальним законом розподілу, хоча добре відомо, що у природі не існує процесів, що підпорядковані цьому закону розподілу (він має місце тільки за граничного переходу, як наприклад, і "дельта-функція").

З огляду на різноманітність випадкових процесів необхідно вирішувати задачі отримання ПВЧ із заданим законом розподілу ймовірності.

Нижче наводиться збірка основних законів розподілу ймовірностей, [], як справочний матеріал для генерування ПВЧ.

2.2.1 Ідентифікація закону розподілу

Найважливішою характеристикою випадкової величини являється її закон розподілу. Знання закону розподілу дозволяє із певною ймовірністю прогнозувати наступне значення випадкової величини, знаходити ймовірність влучення випадкової величини у заданий інтервал, а також моделювати випадкову величину за допомогою генераторів випадкових чисел. Випадкову величину можна замінити у моделі її середнім значенням тільки тоді, коли дослідник упевнений, що це не вплине на результати моделювання. Наприклад, у випадку, коли відхилення випадкової величини від середнього її значення у десятки разів менше за середнє значення, напевно, можна припустити, що випадкова величина є детермінована із значенням, що дорівнює середньому значенню випадкової величини.

Послідовність дій, що виконують для ідентифікації закону розподілу, наведена на рис. 1.

Ідентифікація закону розподі- лу випадкової величини	Формування масиву значень випадкової величини
	Побудова гістограми частот
	Формування гіпотези про вид закону розподілу
	Оцінка значень параметрів закону розподілу
	Перевірка відповідності за критерієм згоди

Рис. 2.1Послідовність дій для ідентифікації закону розподілу випадкової величини

Розглянемо детально кожну дію.

Формування масиву значень, що спостерігаються, випадкової величини. Із попередніх спостережень за випадковою величиною або у ході експериментів отримують певну, бажано велику (більше сотні) кількість значень і запам'ятовують у файлі для проведення подальшого дослідження.

Побудова гістограми частот. Розглянемо, як будується гістограма частот. Спочатку, переглядаючи масив значень випадкової величини ζ , знаходять найменше (min) та найбільше (max) значення випадкової величини, що спостерігається.

Далі, весь інтервал спостережуваних значень випадкової величини від мінімального до максимального значення поділяють на рівні інтервали так, що кожний інтервал має довжину

$$h = \frac{\max - \min}{k}, \quad (8)$$

де k -- кількість інтервалів.

Рекомендована кількість інтервалів k за великої кількості (більше 100) спостережень складає 20. Потім, переглядаючи масив випадкових значень, для кожного i -го інтервалу підраховують кількість влучень випадкової величини у цей інтервал n_i . Гістограма частот представляється прямокутниками, що мають висоту, яка дорівнює кількості влучень n_i , та ширину, що дорівнює довжині інтервалу (рис. 2.2).

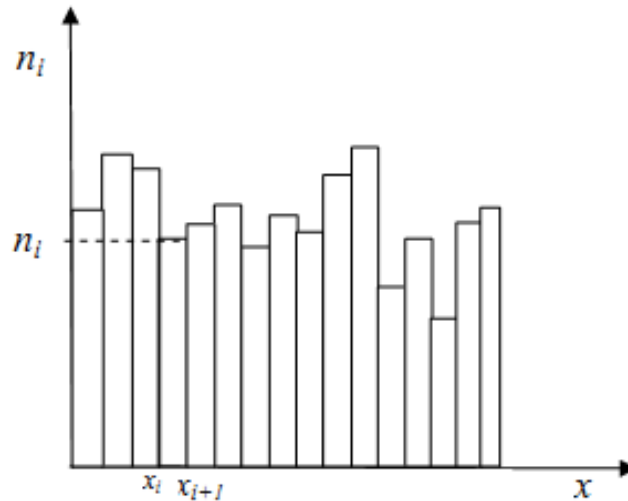


Рис. 2.2 Гістограма частот

За видом гістограми частот роблять припущення про вид закону розподілу. Вид гістограми частот порівнюють із відомими законами розподілу випадкових величин.

2.2.2 Розподіли ймовірностей випадкових величин

Нормальний розподіл

Нормальний розподіл дає гарну модель для реальних явищ, в яких:

- наявна сильна тенденція даних групуватися навколо центру;
- додатні і від'ємні відхилення від центру рівноймовірні;
- частота відхилень швидко зменшується, коли відхилення від центру стають великими.

Параметри розподілу: a, σ ;

Густина розподілу:

$$f(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}};$$

Функція розподілу:

$$F(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt;$$

$$M[x] = a; D[x] = \sigma^2.$$

Коефіцієнт варіації: $V = \sigma/a$; коефіцієнт асиметрії: $A = 0$; коефіцієнт ексцесу: $E = 3$; Мода: $M_0 = a$; Медіана: $M_e = a$.

Для зручності на практиці застосовують нормовану випадкову величину $z = (x - a)/\sigma$, розподіл якої називають стандартним нормальним із нульовим середнім та одиничною дисперсією.

Квантіль -- це аргумент функції розподілу, якій відповідає задана ймовірність p , тобто якщо виразити ймовірність у вигляді

$$p = F(z) = \int_{-\infty}^z f(z)dz,$$

p -квантіль u_p нормально розподіленої випадкової величини зв'язана із квантілю u_p^c , що має стандартний нормальний розподіл, співвідношенням

$$u_p = a + u_p^c \sigma.$$

квантілю буде z , що обчислена із p .

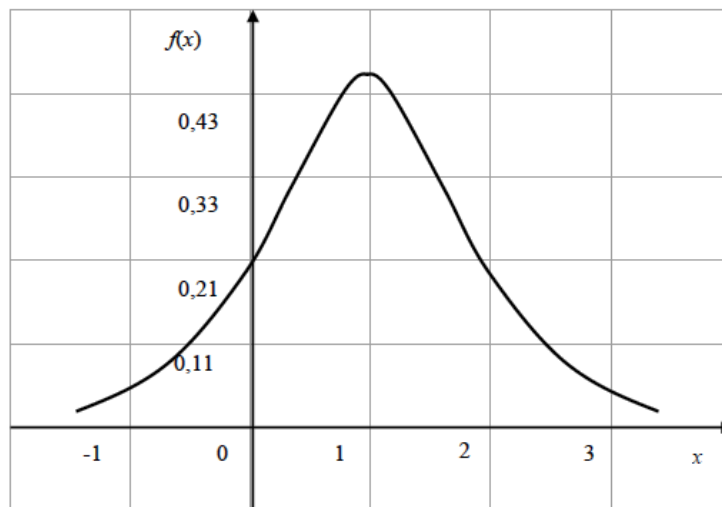


Рис. 2.3 Графік густини нормального розподілу ($\sigma^2 = 1$, $a = 1$)

Логарифмічний нормальний розподіл

Цей розподіл надає можливість описати випадкові величини, логарифм яких розподілений за нормальним законом. Логарифмічний нормальний розподіл -- модель для процесу, що виникає у результаті великої кількості невеликих мультиплікативних похибок.

Густина розподілу:

$$f(x, a, b) = \frac{1}{xb\sqrt{2\pi}} e^{-\frac{(\ln x - a)^2}{2b^2}}; x > 0$$

Функція розподілу:

$$F(x, a, b) = \frac{1}{b\sqrt{2\pi}} \int_0^x e^{-\frac{(\ln t - a)^2}{2b^2}} dt;$$

$$M[x] = e^{a+b^2/2}; D[x] = e^{2a+b^2} (e^{b^2} - 1).$$

$$\nu = \sqrt{e^{b^2} - 1}, \quad A = \sqrt{e^{b^2} - 1} (e^{b^2} + 2);$$

Коефіцієнт ексцесу:

$$E = 3 + (e^{b^2} - 1)(e^{3b^2} + 3e^{2b^2} + 6);$$

$$M_0 = e^{a-b^2}; \quad M_e = e^a.$$

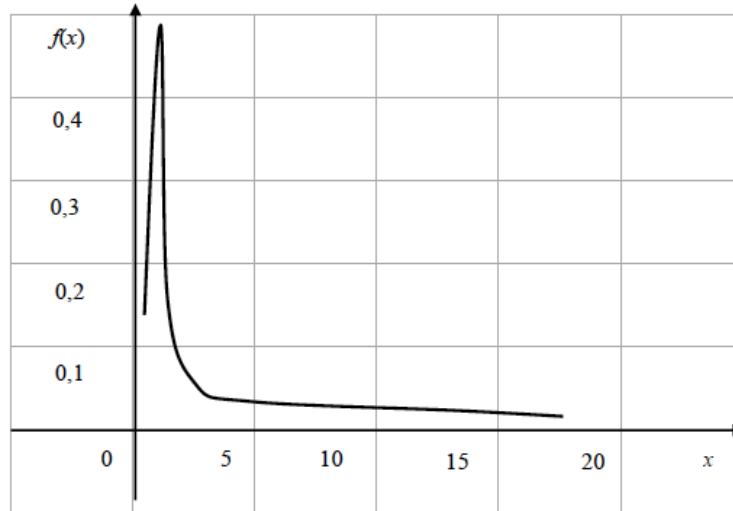


Рис.2.4 Графік густини логонормального розподілу

Гамма-розподіл

Гамма-розподіл є основний розподіл статистики для випадкових величин, що обмежені з одного боку. Гамма-розподіл описує час, необхідний для появи рівно k незалежних подій за умови появи подій із сталою інтенсивністю. Широко використовується у теорії надійності і теорії масового обслуговування.

$$f(x, \alpha, \beta) = \frac{1}{\alpha! \beta^{\alpha+1}} x^{\alpha} e^{-x/\beta}, \quad x > 0, \alpha > 1, \beta > 0;$$

$$F(x, \alpha, \beta) = 1 - \sum_{i=0}^{\alpha} e^{-x/\beta} \left(\frac{x}{\beta} \right)^i \frac{1}{i!}, \quad x > 0, \alpha > 1, \beta > 0;$$

$$M[x] = \beta(\alpha + 1); \quad D[x] = \beta^2(\alpha + 1); \quad \sigma = \sqrt{D[x]}; \quad \nu = 1/\sqrt{\alpha + 1}; \quad A = 2/\sqrt{\alpha + 1};$$

$$E = 3 + 6/(\alpha + 1); \quad M_0 = \alpha\beta.$$

Гамма-розподіл із цілочисловим значенням параметра називається *розподілом Ерланга*. За $\alpha = 0$ гамма-розподіл переходить в *експоненційний закон*.

За $\beta = 2$ гамма-розподіл переходить у *розподіл Релєя*.

Приклад. Випробовуються 4 прилади, інтенсивність відмов яких дорівнює $\lambda = 10^{-5}$ 1/ч. Обчислити ймовірність того, що сумарна наробка приладів не перевищить $3 \cdot 10^5$ годин.

Маємо гамма-розподіл із параметрами $\alpha = 4 - 1 = 3$, $\beta = 1/\lambda = 10^5$.

$$P(x < 300000) = F(300000) = 1 - e^{-3} \left(1 + 3 + \frac{1}{2} 3^2 + \frac{1}{6} 3^3 \right) = 0,352768.$$

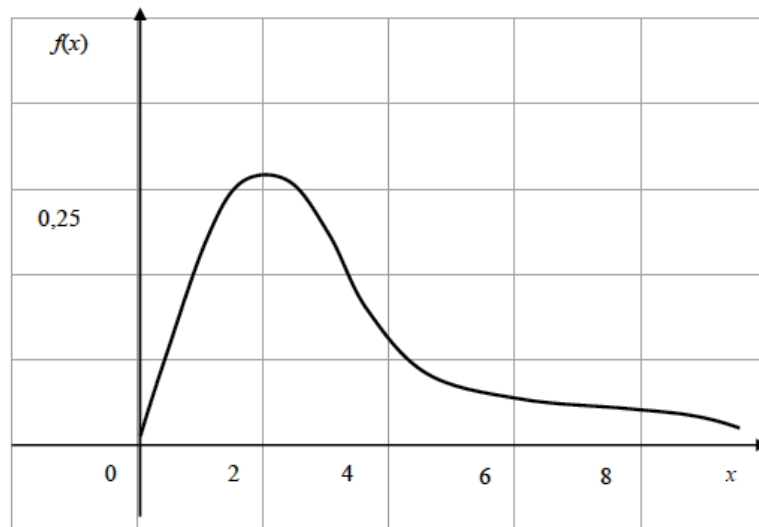


Рис. 2.5 Графік густини Γ -розподілу

Розподіл Вейбулла

Розподіл Вейбулла описує загальний розподіл часу безвідмовної роботи за самих різноманітних інтенсивностях відмов.

$$f(x, \alpha, \beta) = \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-(x/\alpha)^\beta}, \quad x \geq 0, \alpha, \beta > 0;$$

$$F(x, \alpha, \beta) = 1 - e^{-(x/\alpha)^\beta}; \quad x \geq 0, \alpha, \beta > 0;$$

$$M[x] = \mu = \alpha \Gamma(1 + 1/\beta),$$

$$D[x] = \alpha^2 (\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)) = \alpha^2 \Gamma(1 + 2/\beta) - \mu^2; \quad \sigma = \sqrt{D[x]};$$

$$\nu = \sqrt{\frac{\Gamma(1 + 2/\beta)}{\Gamma^2(1 + 1/\beta)}};$$

$$A = \frac{\Gamma\left(1 + \frac{3}{\beta}\right) - 3\Gamma\left(1 + \frac{2}{\beta}\right)\mu + 2\mu^3}{\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]^2};$$

$$E = \frac{\Gamma\left(1 + \frac{4}{\beta}\right) - 4\Gamma\left(1 + \frac{3}{\beta}\right)\mu + 6\Gamma\left(1 + \frac{2}{\beta}\right)\mu^2 - 3\mu^4}{\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \mu^2\right]^2};$$

$$\text{Мода } M_0 = \alpha(1 - 1/\beta)^{1/\beta}; \quad \beta > 1; \quad M_e = \alpha(\ln 2)^{1/\beta}.$$

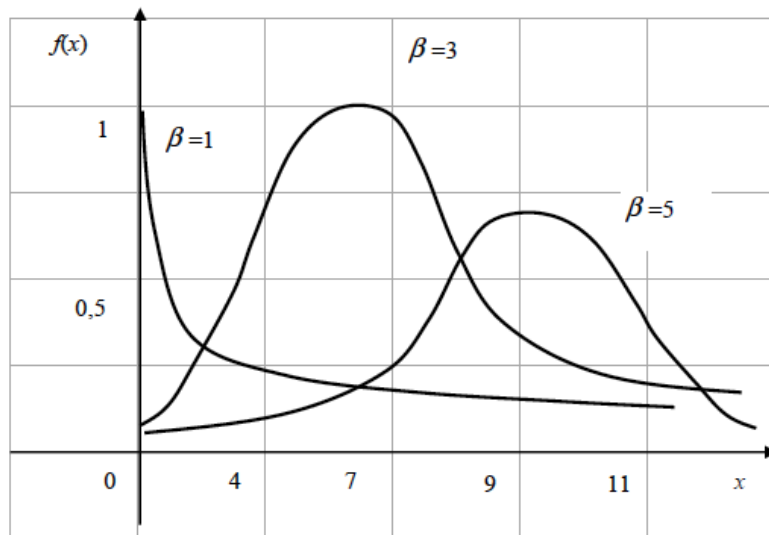


Рис.2.6 Графік густини розподілу Вейбулла

За $\beta = 1$ розподіл Вейбулла переходить в експоненціальний розподіл із параметром α , за $\beta = 2$ -- у розподіл Релея.

Розподіл Релея

Розподіл Релея використовують для опису радикальної похибки на площині, коли похибки за кожною із осей координат незалежні і розподілені за нормальним законом із однаковою дисперсією і нульовим математичним сподіванням.

$$f(x, \alpha) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, \quad x \geq 0, \alpha > 0;$$

$$F(x, \alpha) = 1 - e^{-\frac{x^2}{2\alpha^2}}, \quad x \geq 0, \alpha > 0;$$

$$M[x] = \sqrt{\pi/2}\alpha, \quad D[x] = (2 - \pi/2)\alpha^2;$$

$$\nu = \sqrt{\frac{4}{4-\pi}} = 1,913; \quad A = \frac{2\sqrt{\pi}(\pi-3)}{\sqrt{(4-\pi)^3}} = 0,631; \quad M_0 = \alpha; \quad M_e = 1,177\alpha.$$

Приклад. Випадкова величина розподілена за законом Релея із параметром $\alpha = 4$. Обчислити ймовірність того, що в.в. не перевищить значення $x = 3$ і обчислити 95%-у квантіль розподілу.

Маємо: $P(x < 3) = 1 - e^{-\frac{3^2}{2 \cdot 4^2}} = 0,24516$. Для обчислення 95% квантілі ($x_{0,95}$) використовуємо рівність

$$0,95 = 1 - e^{-\frac{x_{0,95}^2}{2 \cdot 16}} \rightarrow \ln(0,05) = -\frac{x_{0,95}^2}{32} \rightarrow x_{0,95}^2 = 95,86, \rightarrow x_{0,95} \leq 9,971.$$

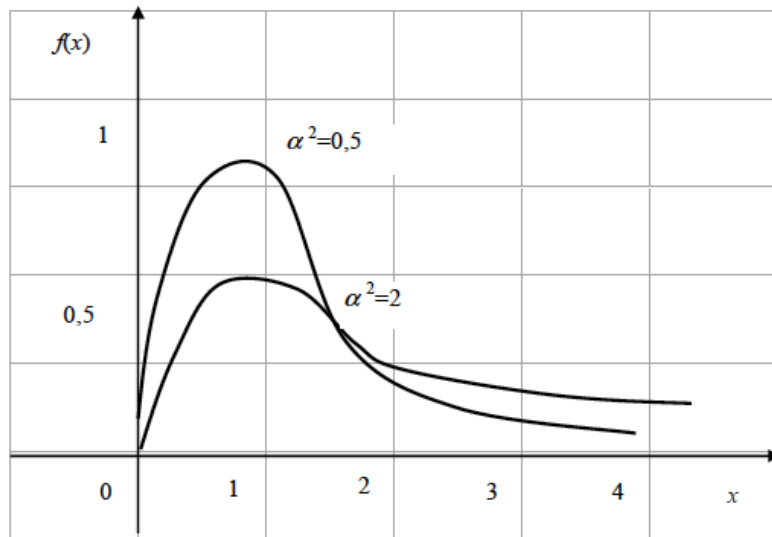


Рис. 2.7 Графік густини розподілу Релєя

Експоненційний розподіл

Описує розподіл часу між незалежними подіями, що з'являються із сталою інтенсивністю: розподіл часу безвідмовної роботи із сталою інтенсивністю.

$$f(x, a) = \frac{1}{a} e^{-x/a} \quad x \geq 0;$$

$$F(x, a) = 1 - e^{-x/a} \quad x \geq 0;$$

$$M[x] = a; \quad D[x] = a^2; \quad \nu = 1; \quad A = 2; \quad E = 9; \quad M_0 = 0; \quad M_e = a \ln 2 = 0,6931;$$

Інтенсивність відмови обчислюється за формулою $\lambda = 1/a$.

Бета-розподіл

Є основний розподіл статистики для випадкових величин, які обмежені з обох боків. Через бета-розподіл можна виразити практично усі розподіли ймовірностей.

$$f(x, \alpha, \beta) = \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} x^\alpha (1-x)^\beta, \quad 0 < x < 1, \alpha, \beta \geq -1;$$

$$F(x, \alpha, \beta) = \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} \int_0^x x^\alpha (1-x)^\beta dx, \quad 0 < x < 1, \alpha, \beta \geq -1;$$

$$M[x] = \frac{\alpha + 1}{\alpha + \beta + 2}; \quad D[x] = \frac{(\alpha + 1)(\beta + 1)}{(\alpha + \beta + 2)^2(\alpha + \beta + 3)};$$

$$\nu = \sqrt{\frac{\beta + 1}{(\alpha + 1)(\alpha + \beta + 3)}};$$

$$A = \frac{2(\beta - \alpha)}{\alpha + \beta + 4} \sqrt{\frac{\alpha + \beta + 3}{(\alpha + 1)(\beta + 1)}};$$

$$E = \frac{3(\alpha + \beta + 2)(\alpha + \beta + 3)}{(\alpha + \beta + 4)(\alpha + 1)(\beta + 1)} \left(\frac{(\alpha + 2)(-\alpha + 2\beta + 1)}{\alpha + \beta + 5} + \frac{(\alpha + 1)(\alpha - \beta)}{\alpha + \beta + 2} \right);$$

$$M_0 = \frac{\alpha}{\alpha + \beta}.$$

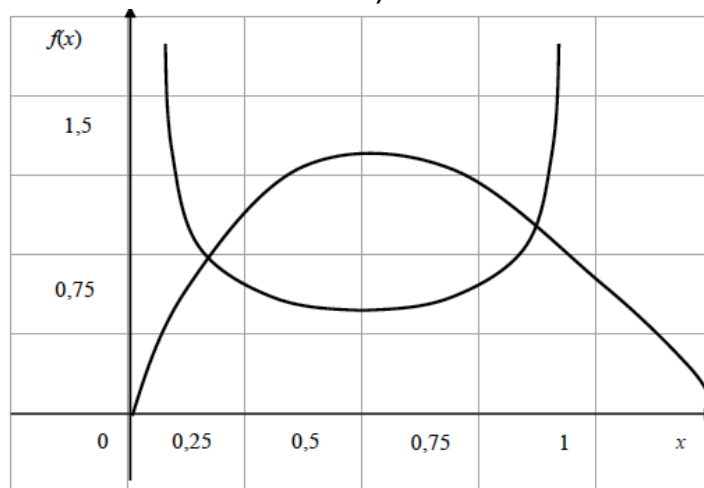


Рис. 2.8 Графік густини β -розподілу

Частинним випадком бета-розподілу є рівномірний розподіл.

Рівномірний розподіл

$$f(x, a, b) = \begin{cases} 1/(b-a), & a \leq x \leq b; \\ 0, & x < a, x > b. \end{cases}$$

$$F(x, a, b) = \begin{cases} 0, & x \leq a; \\ \frac{x-a}{b-a}, & a \leq x \leq b; \\ 1, & x \geq b. \end{cases}$$

$$M[x] = (b+a)/2, \quad D(x) = (b-a)^2/12, \quad \sigma = \sqrt{D(x)}, \quad \nu = \frac{1}{\sqrt{3}} \frac{b-a}{b+a};$$

Коефіцієнт асиметрії: $A = 0$, коефіцієнт ексцесу: $E = 1,8$, $M_e = M[x]$.

Розподіл Пірсона (χ^2)

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2.$$

Цей розподіл виникає як результат додавання квадратів кількох величин, що підпорядковуються нормальному закону із $a = 0$, $\sigma = 1$.

За допомогою χ^2 -розподілу перевіряються гіпотези відносно значень дисперсій, перевіряється узгодження експериментальних даних із теоретичними законами розподілу.

Зміст $f(\chi^2)$ такий самий, що й у нормальному законі: ймовірність того, що величина χ^2 опиняється у заданому інтервалі, дорівнює площі під кривою $f(\chi^2)$. Так, площа під кривою на відрізку від 0 до $n + \sqrt{2n}$ складає $> 90\%$ усієї площі під кривою $f(\chi^2)$. Звідси випливає правило 3σ для закону χ^2 : із ймовірністю $p \geq 0,9$ в.п. $\chi^2 < n + \sqrt{2n}$.

Щоб визначити квантіль χ^2 за заданою ймовірністю та числа ступенів свободи n , необхідно знайти значення у таблиці розподілу χ^2

$$f(\chi^2, f) = \frac{1}{\sqrt{2\Gamma(f/2)}} (\chi^2)^{(f-2)/2} e^{-\chi^2/2}, \quad \chi^2 \geq 0;$$

$$F_f(x) = P(\chi^2(f), x) = \frac{1}{\sqrt{2\Gamma(f/2)}} \int_0^x y^{f/2-1} e^{-y/2} dy, \quad x > 0;$$

$$M[\chi^2(f)] = f; \quad D[\chi^2(f)] = 2f; \quad \nu = \sqrt{2/f};$$

$$A = 2\sqrt{2/f}; \quad E = 3 + 12/f; \quad M_0 = f - 2, \quad f \geq 2; \quad M_e = f - 2/3.$$

f -- число ступенів свободи.

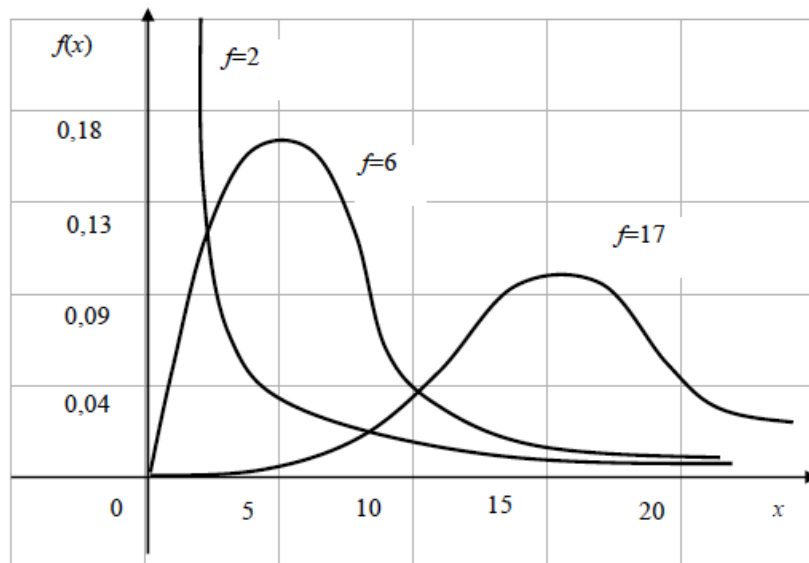


Рис. 2.9 Графік густини розподілу Пірсона

Розподіл Бернуллі (біном. розподіл)

Біноміальному закону підпорядковуються випробування виробів, що не відновлюються, на фіксоване число відмов. Широко застосовується при вибірковому контролю якості продукції та статистичному попереджувальному контролю технологічних процесів.

$$f(x, n, p) = C_n^x p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1;$$

$$F(x, n, p) = \sum_{i=0}^x C_n^i p^i (1-p)^{n-i}, \quad 0 \leq p \leq 1;$$

$$M[x] = np; \quad D[x] = np(1-p); \quad \nu = \sqrt{\frac{1-p}{np}};$$

$$A = (1-2p)\sqrt{n(1-p)}; \quad E = 3 - \frac{6}{n} + \frac{1}{np(1-p)}; \quad M_0 = (n+1)p; \quad M_e = np - 1.$$

Приклад. Ймовірність появи дефектного виробу дорівнює 0,1. Обчислити ймовірність появи у партії із 60 виробів не більше 10 дефектних.

Маємо: $p = 0,1; n = 60; x = 10 \rightarrow np = 6$.

Отримуємо:

$$P(x \leq 10) = F(10; 0,1; 60) = \sum_{i=0}^{10} C_{60}^i \cdot 0,1^i (1-0,1)^{60-i} = 0,9657.$$

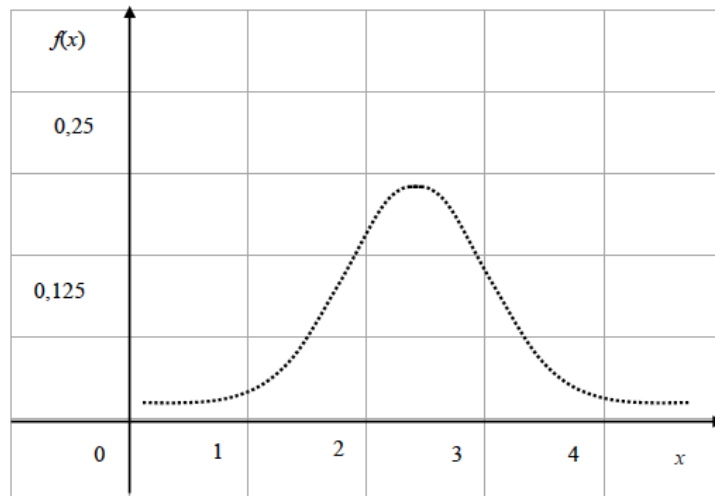


Рис. 2.10 Графік густини розподілу Бернуллі

Розподіл Пуассона

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \lambda > 0;$$

$$F(x, \lambda) = \sum_{i=0}^x \frac{\lambda^i e^{-\lambda}}{i!};$$

$$M[x] = \lambda; \quad D[x] = \lambda; \quad \nu = 1/\sqrt{\lambda}; \quad A = 1/\sqrt{\lambda}; \quad E = 3 + 1/\lambda; \quad M_0 = [\lambda].$$

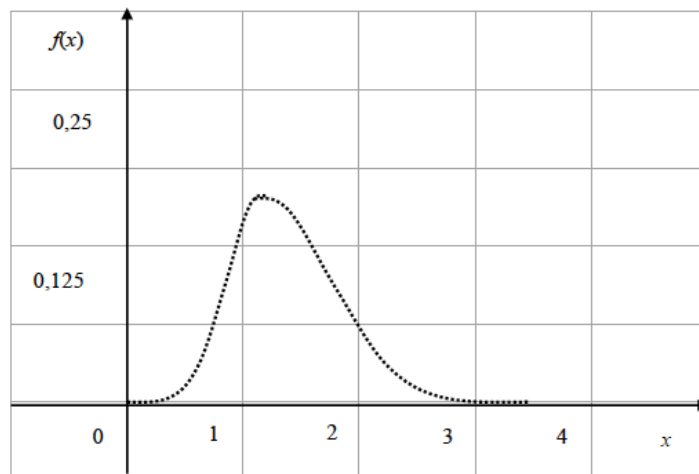


Рис. 2.11 Графік густини розподілу Пуассона

Розподіл Пуассона є гранична форма біноміального розподілу.

Розподіл Ерланга

Функція розподілу

$$A(t) = 1 - \sum_{i=1}^k \frac{(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t}.$$

Густина ерлангівського розподілу:

$$a(t) = A'(t) = \frac{1}{(k-1)!} \lambda (\lambda t)^{k-1} e^{-\lambda t}.$$

$$m = k/\lambda; \quad \sigma = \sqrt{k}/\lambda; \quad \nu = \sqrt{k}/k \leq 1.$$

На рис. (229) представлено графік густини розподілу Ерланга при $\lambda = 0,1$.

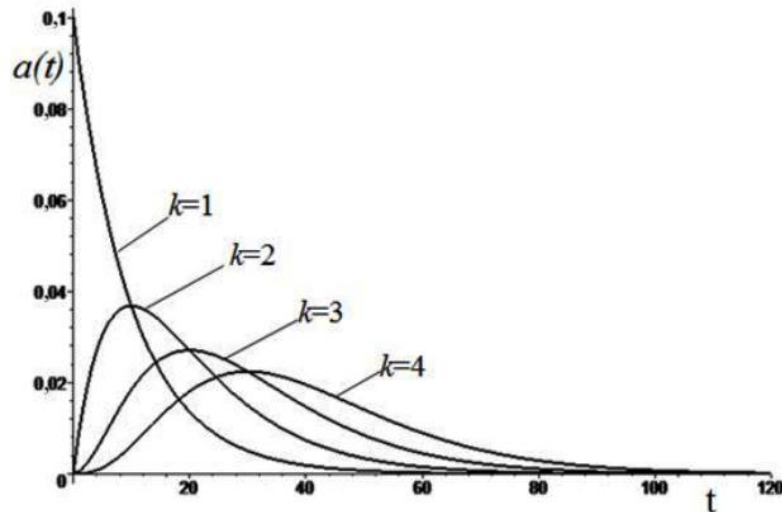


Рис. 2.12 Графік густини розподілу Ерланга

Прийняті такі позначення: $f(x)$ -- густина закону розподілу випадкової величини ζ , μ -- математичне сподівання випадкової величини ζ , σ -- середнє квадратичне відхилення випадкової величини. У таблицях крім графічного зображення закону розподілу наведений зв'язок середнього значення та середнього квадратичного відхилення випадкової величини із параметрами закону.

Оцінка значень параметрів закону розподілу

З формули закону розподілу дослідник робить висновок про параметри закону розподілу, які потрібно визначити. Переважна більшість параметрів закону розподілу може бути оцінена на основі фундаментальних статистичних характеристик випадкової величини таких, як максимальне та мінімальне значення, середнє значення, середнє квадратичне відхилення.

Нагадаємо формули оцінки середнього μ та дисперсії σ випадкової величини ξ :

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad (9)$$

$$\tilde{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \mu)^2, \quad (10)$$

де n -- кількість генерованих випадкових чисел, а хвиляста лінія підкреслює, що розраховується статистична оцінка величини.

У випадку дискретного закону розподілу теоретичне значення частоти влучання

випадкової величини ξ у значення i визначається значенням ймовірності $P(i)$. У випадку неперервного закону розподілу теоретичне значення частоти влучення випадкової величини ξ в i -й інтервал визначається за формулами:

$$p_i^T = \int_{x_{i-1}}^{x_i} f(x) dx = F(x_i) - F(x_{i-1}), \quad (11)$$

де $f(x)$ -- густина закону розподілу випадкової величини ξ , $F(x)$ -- закон розподілу випадкової величини ξ , (x_{i-1}, x_i) -- інтервал. Наприклад, у випадку експоненціального закону розподілу теоретичне значення частоти влучання в інтервал може бути визначене за формулою:

$$p_i^T = \int_{x_i}^{x_{i+1}} \lambda e^{-\lambda x} dx = (e^{-\lambda x_i} - e^{-\lambda x_{i+1}}), \quad (12)$$

де параметр λ має бути оцінений за формулою

$$\lambda = \frac{1}{\tilde{\mu}} = \frac{1}{\tilde{\sigma}}. \quad (13)$$

А у випадку закону розподілу Пуассона -- за формулою:

$$p_i^T = \frac{\lambda^i}{i!} e^{-\lambda}, \quad (14)$$

де параметр λ оцінюється за формулою:

$$\lambda = \tilde{\mu} = \tilde{\sigma}. \quad (15)$$

Перевірка відповідності випадкових чисел обраному закону розподілу. Відповідність обраного закону розподілу заданим випадковим числам (при кількості спостережуваних чисел більше ста) перевіряється за критерієм згоди χ^2 :

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i^T)^2}{np_i^T}, \quad (16)$$

де n_i -- кількість влучань в i -й інтервал, що спостерігається, np_i^T -- кількість влучань в i -й інтервал, що очікується за теоретичним законом розподілу. З формули (16) видно, що основною ідеєю критерію χ^2 є вимірювання розбіжності між кількістю влучань в i -й інтервал, що спостерігається, та очікується за теоретичним законом розподілу (рис.2.6.).

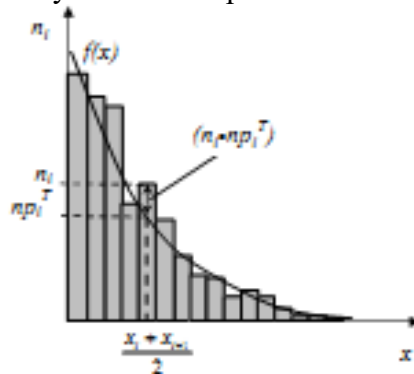


Рис. 2.13 До пояснення змісту критерію χ^2

Розраховане значення χ^2 порівнюється з табличним значенням критерію χ_{kr}^2 , яке взяте при рівні значимості $\alpha = 0,05$ та кількості ступенів свободи, рівній кількості інтервалів у гістограмі частот k мінус 1 мінус кількість параметрів закону розподілу.

Якщо $\chi^2 < \chi_{kr}^2$, із довірчою ймовірністю 0,95 можна стверджувати, що знайдений закон розподілу відповідає значенням випадкової величини ξ , що спостерігається. Інакше потрібно змінити параметри розподілу або припустити інший закон розподілу.

Довірча ймовірність результату означає, що якщо, наприклад, за допомогою критерію χ^2 з'ясували відповідність закону випадкових чисел закону розподілу, то гарантувати цю відповідність можемо не більше як із ймовірністю 0,95. Із ймовірністю 0,05 могло статися, що спостережувані значення випадково виявилися не відповідними закону розподілу. Довірча ймовірність означає також, що, якщо провести повторний експеримент із тією самою випадковою величиною, то із ймовірністю 0,95 результат ідентифікації закону розподілу співпаде із отриманим попереднім результатом.

Критерій χ^2 має такі обмеження для застосування: по-перше, кількість випадкових чисел має бути не менша за 100; по-друге, кількість влучань у кожний інтервал має бути більшою за 5:

$$n \geq 100, \quad n_i \geq 5, \quad i = \overline{1, k}. \quad (17)$$

Якщо кількість влучень в якийсь інтервал менша за 5, то потрібно об'єднати сусідні інтервали і домогтися, щоб кількість влучень стала більшою або рівною 5. Допускається об'єднання як двох інтервалів, так і більше.

Якщо кількість досліджуваних чисел надто мала (менше ста), то потрібно застосовувати *критерій Колмогорова--Смірнова* (або, як його ще називають, λ -критерій). За критерієм Колмогорова--Смірнова відповідність випадкових чисел, що спостерігаються, обраному закону розподілу оцінюється за значенням найбільшої абсолютної різниці []:

$$\lambda = \sqrt{n} \max_i |w_i - w_i^T|, \quad w_i = \sum_{j=1}^i \frac{n_j}{n}, \quad w_i^T = F(x_i), \quad (18)$$

де w_i -- спостережувані значення сумарної частоти влучань в i -й інтервал, n_j -- кількість влучань випадкової величини в j -й інтервал, n -- кількість випадкових чисел, що спостерігаються, w_i^T -- теоретичні значення сумарної частоти влучань в i -й інтервал, розраховані за законом розподілу $F(x)$ випадкової величини ζ , n_i -- середина i -го інтервалу.

З формули (18) випливає, що на відміну від χ^2 -критерія λ -критерій порівнює сумарні частоти влучення в i -й інтервал (рис.2.7).

Розраховане значення λ -критерію порівнюють з табличним значенням критерію, яке взяте при рівні значимості $\alpha = 0,05$. Якщо $\lambda < \lambda_{kr}$, із довірчою ймовірністю 0,95 можна стверджувати, що випадкова величина ζ має закон розподілу $F(x)$.

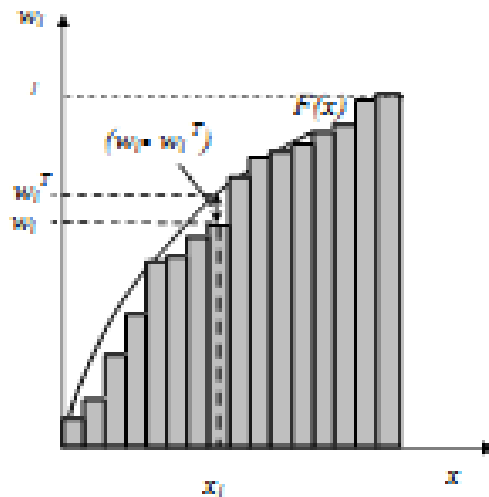


Рис.2.14 До пояснення змісту λ -критерію (Колмогорова--Смірнова)

В [5] зауважується, що критерій Колмогорова -- Смірнова вимагає, щоб оцінка параметрів розподілу виконувалась за даними, які не використовувались для розрахунку λ -критерію.

2.2.3 Апроксимація функціональної залежності

Будь-який елемент системи описується залежністю множини вихідних його змінних від множини вхідних. У загальному випадку така залежність може бути представлена функціональною залежністю. Відновлення функціональної залежності змінних моделі за результатами спостережень називають апроксимацією. Заміна об'єкта, який є елемент системи, його функціональною залежністю є, певною мірою, спрощення. Причини, з яких дослідник змушений використовувати апроксимацію різні. Можливо, об'єкт надто складний, щоб представити його функціонування детальніше. Можливо, спрощене представлення об'єкта дозволить спростити загальну модель системи.

Сформулюємо задачу апроксимації. Розглядається об'єкт, що має одну вихідну змінну y та кілька вхідних змінних, представлених вектором $x = (x_1, x_2, \dots, x_m)$.

Відомі дані спостережень про значення вхідної та вихідної змінної:

Дані спостережень

Змінні	Спостереження			
	1	2	...	n
Вхідна змінна x_1	x_{11}	x_{12}	...	x_{1n}
Вхідна змінна x_2	x_{21}	x_{22}	...	x_{2n}
...				
Вхідна змінна x_m	x_{m1}	x_{m2}	...	x_{mn}
Вихідна змінна y	y_1	y_2	...	y_n

Рис. 2.15 Дані спостережень

Задача апроксимації полягає у вішуванні функціональної залежності заданого виду, що найліпше відповідає даним спостережень з точки зору критерію найменших квадратів. Нехай y_i -- значення вихідної змінної, що спостерігається, $y_i^f = f(x_i)$ -- розраховані за

функціональною залежністю $f(x_i)$ значення вихідної змінної. Ідея критерію найменших квадратів полягає у тому, що розбіжність між значеннями вихідної змінної, що спостерігається, і розраховується, має бути мінімальною. Якщо розбіжність вимірювати сумою різниць $(y_i - y_i^f)$, може статися так, що навіть за великих, але різних за знаком, різниць загальна розбіжність має мале значення. Тому розбіжність вимірюють сумою квадратів різниць $(y_i - y_i^f)^2$. Отже, критерій найменших квадратів має вигляд:

$$F = \sum_i (y_i^f - y_i)^2 = \sum_i [f(x_i) - y_i]^2 \rightarrow \min, \quad (19)$$

де мінімум у загальному випадку розглядається по відношенню до різних пропонованих функцій $f(x_i)$.

Апроксимацію функціональної залежності виконують у такий спосіб:

- 1) формують масив значень, що спостерігаються;
- 2) формують припущення про вид математичної функції -- параметри функціональної залежності;
- 3) відшуковують значення параметрів функціональної залежності b за критерієм найменших квадратів;
- 4) оцінюють якість знайденої функціональної залежності методами багатofакторного кореляційно - регресійного аналізу.

Формування масиву значень, що спостерігаються.

На відміну від функції однієї змінної, задача апроксимації функції багатьох змінних має розв'язок тільки за умови, що змінні $x_1, x_2, \dots, x_m$ не залежать одна від одної. Дійсно, якщо, наприклад, $x_1 = g(x_2)$, змінна x_2 має бути викреслена зі списку змінних. Якщо дослідник помилково включить залежну змінну у список змінних, параметри функціональної залежності не можуть бути знайдені.

Формування гіпотези про вигляд функціональної залежності.

Застосування методу найменших квадратів вимагає, щоб функціональна залежність, яка відшукується, мала лінійний відносно параметрів b вигляд:

$$f(x, b) = b_0 + b_1 f_1(x) + b_2 f_2(x) + \dots + b_k f_k(x), \quad (20)$$

де $x = (x_1, x_2, \dots, x_m)$ -- вектор змінних.

Припущення про вигляд функцій $f_i(x)$ дослідник має зробити, ґрунтуючись на своїх знаннях та досвіді. Параметри b функціональної залежності $f(x, b)$ вибираються так, щоб забезпечити мінімум критерію найменших квадратів.

Оцінка значень параметрів функціональної залежності.

Значення критерію найменших квадратів визначається за формулою:

$$F(b) = \sum_{i=0}^n [f(x_i, b) - y_i]^2 = \sum_{i=0}^n [b_0 + b_1 f_1(x_i) + b_2 f_2(x_i) + \dots + b_k f_k(x_i)]^2 \rightarrow \min, \quad (21)$$

де $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, y_i -- дані досліджень про об'єкт.

2.2.4 Кореляційно-регресійний аналіз функціональної залежності

За малого обсягу даних виникає необхідність перевірки отриманої функціональної залежності.

Для цього використовують методи багатofакторного кореляційно- регресійного

аналізу.

Оцінку практичної значущості отриманої функціональної залежності проводять за допомогою індексу кореляції, який характеризує щільність зв'язку:

$$R = \sqrt{\sigma_{\hat{y}}^2 / \sigma_{\text{ccag}}^2}, \quad (22)$$

де

$$\sigma_{\hat{y}}^2 = \frac{\sum [f(x, b) - \bar{y}]^2}{n-1}$$

-- факторна дисперсія результативної ознаки y ,

$$\sigma_{\text{ccag}}^2 = \frac{\sum (y - \bar{y})^2}{n-1}$$

-- його загальна дисперсія.

Індекс детермінації R^2 характеризує, яка частина загальної варіації результативної ознаки y пояснюється фактором x . Якщо більше половини загальної варіації результативної ознаки y пояснюється впливом фактора x , застосування методів функціонального аналізу для вивчення кореляційного зв'язку є виправдані, а синтезовані при цьому аналітичні моделі визнаються придатними для їх практичного застосування. Звідси, функціональна залежність практично значима, якщо $R^2 > 0,5$.

Оцінку істотності індексу кореляції проводять за критерієм Фішера:

$$F = \frac{R^2}{1-R^2} \frac{n-k-1}{k}. \quad (23)$$

Знайдене значення критерію порівнюють із критичним значенням критерію Фішера, який беруть із статистичних таблиць при рівні значимості 0,05, кількості ступенів свободи чисельника $n-k-1$ та кількості ступенів свободи знаменника k . Якщо $F > F_{kr}$, значення індексу кореляції визнається істотним.

Оцінку статистичної значимості коефіцієнтів регресії виконують за критерієм Ст'юдента у такий спосіб. Спочатку розраховують діагональні елементи матриці

$$d_j = [(X^T X)^{-1}]_{jj}, \quad j = \overline{0, k}. \quad (24)$$

Потім знаходять дисперсію

$$s^2 = \frac{\sum [f(x, b) - y_i]^2}{n-k-1}.$$

Значення критерію Ст'юдента, що спостерігається, відшукується окремо для кожного параметра b_j за формулою:

$$t_j = \frac{|b_j|}{\sqrt{d_j s^2}}. \quad (25)$$

Критичне значення критерію Ст'юдента беруть із статистичних таблиць за рівня значимості 0,05 та кількості ступенів свободи $n-k-1$. Якщо $t_j > t_{kr}$, знайденого із таблиць критерію Ст'юдента за рівня значимості $\alpha = 0,05$ та кількості ступенів свободи $n-2$, коефіцієнт b_j визнається значимим.

Довірчий інтервал для коефіцієнту b_j визначається за формулою:

$$b_j \pm t_{kr} \sqrt{d_j s^2}. \quad (26)$$

Знання довірчого інтервалу надає досліднику можливість оцінити величину похибки отриманого значення параметру.

2.2.5 Апроксимація залежності за методом χ^2

У методі апроксимації за методом найменших квадратів вважається, що точність значень, що спостерігаються, приблизно однакова. У випадках, коли відомо, що точність спостережуваних значень суттєво відрізняється, доцільно використовувати апроксимацію за методом χ^2 .

Оцінка значень параметрів функціональної залежності здійснюється за методом χ^2 :

$$\chi^2(b) = \sum_{i=0}^n \left(\frac{f(x, b) - y_i}{\sigma_i} \right)^2 \rightarrow \min, \quad (27)$$

де σ_i -- середнє квадратичне відхилення значення y_i , що спостерігається.

У випадку, коли точність визначення всіх y_i однакова, отримуємо таку формулу визначення критерію χ^2 :

$$\chi^2(b) = \sum_{i=0}^n \left(\frac{f(x, b) - y_i}{\sigma_i} \right)^2 = \frac{1}{\sigma^2} \sum_{i=0}^n (f(x, b) - y_i)^2 \rightarrow \min, \quad (28)$$

де σ -- однакове для всіх значень y_i , що спостерігаються, середнє квадратичне відхилення.

Розраховане значення критерію χ^2 порівнюють із табличним значенням χ_{kr}^2 , яке знаходять при кількості ступенів свободи, що дорівнює кількості даних, що спостерігаються, мінус один і мінус кількість шуканих параметрів функціональної залежності, та рівні довірчої ймовірності 0,95.

Якщо $\chi^2 < \chi_{kr}^2$, функціональна залежність, що знайдена за методом χ^2 , із довірчою ймовірністю 0,95 відповідає даним спостережень. В іншому випадку потрібно змінити вид функціональної залежності.

Використання критерію χ^2 має суттєві переваги порівняно з використанням критерію найменших квадратів. Проте використання його обмежене, оскільки дослідник не завжди має інформацію про точність вимірювання даних.

2.2.6 Приклади розв'язання задач

Задача 1. За даними спостережень знайти апроксимуючу функцію, що має вигляд $y = b_0 + b_1 x_1^2 + b_2 x_2$, та оцінити її якість.

Дані спостережень

$$\begin{bmatrix} X_1 & 4 & 3 & 4 & 4 & 2 & 5 & 4 & 3 & 5 & 4 & 3 & 5 & 3 & 5 & 4 & 3 & 2 \\ X_2 & 1 & 15 & 13 & 11 & 13 & 13 & 15 & 14 & 15 & 11 & 12 & 12 & 11 & 12 & 14 & 11 & 11 \\ Y & 40 & 15 & 38 & 40 & 2 & 65 & 36 & 16 & 63 & 40 & 18 & 66 & 19 & 63 & 37 & 19 & 4 \end{bmatrix}.$$

Розв'язання. За даними спостережень із урахуванням функціональної залежності складемо матрицю X та вектор y , розрахуємо параметри функціональної залежності b :

$$b = (X^T X)^{-1} X^T Y = \begin{bmatrix} -4.915 \\ 2.934 \\ -0.323 \end{bmatrix}.$$

Отже, знайдена така функціональна залежність:

$$y_{\text{reg}} = f(x) = -4.915 + 2.934x_1^2 - 0.323x_2.$$

Оцінимо якість знайденої функціональної залежності. Розрахуємо критерій найменших квадратів за формулою (21).

Розрахунок критерію найменших квадратів

$$\begin{bmatrix} x_1 & 4 & 3 & 4 & 4 & 2 & 5 & 4 & 3 & 5 \\ & 4 & 3 & 5 & 3 & 5 & 4 & 3 & 2 & \\ x_2 & 1 & 15 & 13 & 11 & 13 & 13 & 15 & 14 & 15 \\ & 11 & 12 & 12 & 11 & 12 & 14 & 11 & 11 & \\ y & 40 & 15 & 38 & 40 & 2 & 65 & 36 & 16 & 63 \\ & 40 & 18 & 66 & 19 & 63 & 37 & 19 & 4 & \\ y_{\text{reg}} & 41,7 & 16,6 & 37,8 & 38,5 & 2,6 & 64,2 & 37,2 & 17,0 & 63,6 \\ & 38,5 & 17,6 & 64,5 & 17,9 & 64,5 & 37,5 & 17,9 & 3,3 & \\ |y_{\text{reg}} - y| & 1,7 & 1,6 & 0,2 & 1,5 & 0,6 & 0,8 & 1,2 & 1,0 & 0,6 \\ & 1,5 & 0,4 & 1,5 & 1,1 & 1,5 & 0,5 & 1,1 & 0,7 & \\ (y_{\text{reg}} - y)^2 & 2,89 & 2,69 & 0,03 & 2,34 & 0,38 & 0,60 & 1,38 & 0,93 & 0,34 \\ & 2,34 & 0,15 & 2,11 & 1,14 & 2,40 & 0,25 & 1,14 & 0,54 & \end{bmatrix}.$$

$$\sum (y_{\text{reg}} - y)^2 = 21.65.$$

Отже, значення критерію найменших квадратів досягає значення 21,65, що порівняно із значеннями вихідної величини є досить прийнятне значення. Із наведених розрахунків видно також, що відхилення розрахованого за функціональною залежністю значення від спостережуваного значення не перевищує 1,7, що не перевищує 5% значення вихідної величини.

Проведемо кореляційно-регресійний аналіз функціональної залежності. Обчислимо індекс кореляції:

$$\bar{Y} = \frac{\sum y_i}{17} = 34.2; \quad \sigma_{\text{fact}} = \frac{(y_{\text{reg}} - 34.2)^2}{16} = 443.5,$$

Отже, за шкалою Чеддока зв'язок оцінюється як дуже сильний. Значення індексу детермінації свідчить про те, що 99,7% варіації вихідної змінної y пояснюється змінюванням значень вхідних змінних x_1 та x_2 .

Обчислимо довірчі інтервали за формулами (19)—(26):

$$d = \begin{bmatrix} 1.16 \\ 0.001 \\ 0.006 \end{bmatrix}; \quad s^2 = \frac{(y_{\text{reg}} - y_i)^2}{17 - 3 - 1} = 1.67; \quad t = \begin{bmatrix} 3.5 \\ 65.3 \\ 3.2 \end{bmatrix};$$

$$t_{kr} = qt(0.95, 17 - 3 - 1) = 1.77, \quad t_{kr} \sqrt{d_j s^2} = 1.77 \sqrt{1.67 d} = \begin{bmatrix} 2.46 \\ 0.08 \\ 0.18 \end{bmatrix}.$$

Отже, довірчі інтервали параметрів функціональної залежності задаються наступними значеннями:

$$b_{\min} = \begin{bmatrix} -7.37 \\ 2.85 \\ -0.50 \end{bmatrix}; \quad b_{\max} = \begin{bmatrix} -2.46 \\ 3.01 \\ -0.14 \end{bmatrix}.$$

Відповідь. Дані спостережень відповідають функціональній залежності

$$f(x) = -4.915 + 2.934x_1^2 - 0.3323x_2$$

із значенням критерію найменших квадратів 21,65.

Результати кореляційно-регресійного аналізу свідчать, що зв'язок між змінними y та x дуже сильний, усі параметри функціональної залежності являються значимими. Довірчі інтервали свідчать про те, що найбільш неточно знайдений перший параметр функціональної залежності ($\pm 2,46$), а найбільш точно знайдений другий параметр функціональної залежності ($\pm 0,08$).

Задача 2. Інтенсивність надходження покупців до маркету впродовж дня змінюється. Знайти апроксимуючу функцію, що наближено моделює інтенсивність надходження покупців залежно від часу, та визначити якість знайденої функції.

Дані інтенсивності надходження покупців до маркету впродовж дня Дані інтенсивності надходження

t	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Y	15	23	16	9	11	16	17	23	32	41	53	28	12	8

Розв'язання. Оскільки спостерігається два піки збільшення інтенсивності надходження покупців (на 9-ту та 18-ту годину), апроксимуюча функція має бути принаймні 4-ого порядку:

$$y = b_0 + b_1t + b_2t^2 + b_3t^3 + b_4t^4.$$

Побудуємо матрицю для відшукування параметрів функціональної залежності і розрахуємо параметри функціональної залежності:

$$b = (X^T X)^{-1} X^T Y = \begin{bmatrix} 362.3 \\ -86.5 \\ 7.02 \\ -0.2 \\ -0.001 \end{bmatrix}.$$

Отже, апроксимуюча функціональна залежність має вигляд:

$$y_{\text{reg}} = f(x) = -362.3 - 86.6t + 7.02t^2 - 0.2t^3 - 0.001t^4.$$

Оцінимо якість знайденої функціональної залежності. По-перше, розрахуємо критерій найменших квадратів за формулою (21):

$$F(b) = \sum_{i=0}^n (f(x_i, b) - y_i)^2 = 873.6.$$

Значення критерію найменших квадратів близько 30^2 .

По-друге, оцінимо якість знайденої функціональної залежності. Розрахуємо індекс кореляції за формулою (22):

$$\bar{y} = \frac{\sum y_i}{15} = 20.6; \quad \sigma_{\text{fact}} = \frac{\sum (y_{\text{reg}_i} - 20.6)^2}{14} = 112.7; \quad \sigma_a = \frac{\sum (y_i - \bar{y})^2}{14} = 175.1;$$

$$R = \sqrt{\sigma_{\text{fact}} / \sigma_a} = 0.8 \rightarrow R^2 = 0.64.$$

Отже, за шкалою Чеддока зв'язок оцінюється як сильний, але не дуже сильний. Значення індексу детермінації свідчить про те, що 64% змінювання інтенсивності надходження покупців пояснюється змінюванням часу.

Оцінимо значимість параметрів функціональної залежності (24), (25):

$$d = \begin{bmatrix} 2286 \\ 199.4 \\ 2.27 \\ 0.005 \\ 0.1 \cdot 10^{-5} \end{bmatrix}; s^2 = \frac{\sum_i (y_{\text{reg}_i} - y_i)^2}{15 - 4 - 1} = 87.4;$$

$$d = \begin{bmatrix} 5.9 \\ 4.7 \\ 3.6 \\ 2.2 \\ 0.8 \end{bmatrix}; t_{kr} = qt(0.95, 15 - 4 - 1) = 1.8.$$

Порівнюючи розраховані значення t_j з критичним значенням t_{kr} , знаходимо, що значення параметра b_4 є незначиме.

Розрахуємо точність визначення параметрів функціональної залежності, які є значимі, за формулою (26):

$$1.8\sqrt{87.36d} = \begin{bmatrix} 817.8 \\ 241.5 \\ 25.8 \\ 1.2 \end{bmatrix}.$$

Відповідь. Дані спостережень відповідають функціональній залежності

$$f(x) = -362.3 - 86.5t + 7.02t^2 - 0.2t^3 - 0.0001t^4$$

із значенням критерію найменших квадратів 873,6.

Результати кореляційно-регресійного аналізу свідчать, що зв'язок між змінними y та x сильний. Усі параметри функціональної залежності, окрім останнього, являються значимими. Точність визначення довірчих інтервалів першого та другого параметрів функціональної залежності недостатня, тому рекомендується збільшити кількість даних спостережень.

2.3 Генератори ПВЧ

2.3.1 Обчислення у полях Галуа

Групою G називається множина елементів $(a_1, a_2, \dots, a_n) \in G$ (із заданою асоціативною бінарною операцією), для яких виконуються такі умови:

1. $a \circ e = a$ -- існування деякого одиничного елемента $e \in G$;

2. $a \circ \bar{a} = e$ -- існування зворотного елемента $\bar{a} \in G$ для будь-якого елемента групи $a \in G$.

Кількість елементів у групі називається порядком групи.

Множина цілих чисел $(\dots, -3, -2, -1, 0, 1, 2, 3, \dots) \in \mathbb{Z}$ відносно бінарної операції $''+''$ (арифметичне додавання) утворює групу, оскільки у неї елемент $e = 0$ відіграє роль одиничного (тобто $a + e = a + 0 = a$) і для будь-якого числа $a \in \mathbb{Z}$ можна знайти таке, що їхня сума $e = 0$. Наприклад, для $a = 4$, зворотним буде $b = -4$, тоді $a + b = 4 + (-4) = e$. Описана множина далі буде називатися *адитивною групою*.

Множина тих самих цілих чисел $(\dots, -3, -2, -1, 0, 1, 2, 3, \dots) \in \mathbb{Z}$ відносно бінарної операції $''\cdot''$ (множення) не утворює групу, оскільки у ній не можна знайти зворотного елемента для довільного $a \in \mathbb{Z}$. Вочевидь, роль одиничного елемента тут буде відігравати власне одиниця $e = 1$ (тобто $a \cdot e = a \cdot 1 = a$), але для відшукування зворотного елемента необхідно для довільного елемента $a \in \mathbb{Z}$ знайти таке b , щоб виконувалась рівність $a \cdot b = e$. Наприклад, для $a = 5$ ми маємо записати $b = 1/5 = 0,2$ і тоді $5 \cdot 0,2 = 1$. Але $b = 1/5 = 0,2 \notin \mathbb{Z}$, тобто зворотне число не є ціле і не належить множині $\mathbb{Z}$. Така множина елементів називається *півгрупою* (мультиплікативною півгрупою).

Приклад 1. Класичний стрілочний годинник має на циферблаті значення $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$.

Якщо заняття почалися об 11 годині і будуть тривати 2 години, то коли вони закінчатся? Зазвичай відповідь обчислюється так: $11 + 2 = 13$, але числа 13 годиннику не намальовано. Тоді ми віднімаємо $13 - 12 = 1$ і кажемо, що заняття закінчилися в 1 годину.

Приклад 2. Якщо ми лягли спати о 10-й годині і проспали 9 годин, то коли ми прокинулися? Обчисливши $10 + 9 = 19 = 19 - 12 = 7$, ми кажемо, що прокинулися о 7 годині.

Для того, щоб позначити, що для нас $19 \equiv 7$, ми використали інший знак рівності $''\equiv''$.

Приклад 3. Поїзд № 240Э Москва -- Владивосток ходить кожний день із Ярославського вокзалу. Відправлення о 20:00. Час у дорозі 6 днів 15 годин. Коли (за Московським часом) поїзд прибуде у Владивосток? Обчислюючи, отримаємо $20 + 6 \cdot 24 + 15 = 179$. Тепер розділимо 179 на 12 і знайдемо залишок $179/12 = 14 + 11/12$ або $179 = 14 \cdot 12 + 11$. Тобто поїзд прибуде у Владивосток об 11 годині. Для визначення даного часу ми обчислили залишок від ділення 179 на 11. Цей залишок і є відповідь. Математично даний факт записується так $179 = 11(\text{mod}12)$.

Тепер розв'язки попередніх задач записуються у вигляді: $11 + 2 = 13 = 1(\text{mod}12)$, $10 + 9 = 19 = 7(\text{mod}12)$.

Кільцем називається множина елементів із двома бінарними операціями ($''+''$, $''\cdot''$), яка є адитивна група, але мультиплікативна півгрупа. Розглянемо множину із трьох чисел $(0, 1, 2)$ і складемо для неї таблицю додавання і таблицю множення за модулем 3.

Дана множина утворює кільце і позначається як $\mathbb{Z}_3$.

Для кільця $\mathbb{Z}_4$ таблиці додавання і множення за $(\text{mod}4)$ мають такий вигляд

Таблиця 2.2. Таблиці додавання і множення за $(\text{mod}4)$

+	0	1	2	3	×	0	1	2	3
0	0	1	2	3	0	0	0	0	0
1	1	2	3	0	1	0	1	2	3
2	2	3	0	1	2	0	2	0	2
3	3	0	1	2	3	0	3	2	1

Зазначимо, що для елемента 2 не існує мультиплікативного зворотного. Тобто у таблиці множення немає такого числа x , яке дало б $2 \cdot x = 1$.

Для кільця Z_5 таблиці додавання і множення за $(\text{mod } 5)$ мають вигляд

Табл.2.3 Таблиці додавання і множення за $(\text{mod } 5)$

+	0	1	2	3	4	$\times$	0	1	2	3	4
0	0	1	2	3	4	0	0	0	0	0	0
1	1	2	3	4	0	1	0	1	2	3	4
2	2	3	4	0	1	2	0	2	4	1	3
3	3	4	0	1	2	3	0	3	1	4	2
4	4	0	1	2	3	4	0	4	3	2	1

Із таблиці множення для Z_5 можна бачити, що для кожного ненульового елемента існує зворотний: $1 \cdot 1 = 1$, $2 \cdot 3 = 1$, $3 \cdot 2 = 1$, $4 \cdot 4 = 1$. Тобто для 2 зворотним елементом буде 3, для 3 -- це 2, для 1 -- це 1, а для 4 -- це 4. Виходить, що у кільці Z_5 елементи (0, 1, 2, 3, 4) утворюють групу як за додаванням, так і за множенням.

Полем називається множина елементів із двома бінарними операціями ("+" і "×"), яке є групою як за додаванням, так і за множенням (тобто адитивною і мультиплікативною групою одночасно).

Неважко перевірити, що усі кільця Z_p , для яких p є просте число, є скінченні поля.

Для складання таблиць множення у скінченних полях і виконання у них арифметичних операцій часто користуються таблицями індексів. Таблиця індексів для Z_7 будується таким чином. Візьмемо довільне число (наприклад, 2) і шукатимемо залишки при діленні його степеню на 7:

$$2^0 \equiv 1(\text{mod } 7), 2^1 \equiv 2(\text{mod } 7), 2^2 \equiv 4(\text{mod } 7), 2^3 \equiv 1(\text{mod } 7), \dots$$

Зведемо відповіді у таблицю 2.4.

Табл. 2.4. Таблиця індексів для Z_7

n	0	1	2	3	4	5	6
2^n	1	2	4	1	2	4	1

Із таблиць 2.3, 2.4 видно, що значення залишків $2^n(\text{mod } 7) = (1, 2, 4)$ не приймають усіх чисел кільця (1, 2, 3, 4, 5, 6), тобто 2 не є генератор Z_7 . Тепер візьмемо в якості основи число 3: Таблиця індексів для Z_7 з основою 3

n	0	1	2	3	4	5	6
3^n	1	3	2	6	4	5	1

Із таблиці (експонент) видно, що $3^n(\text{mod } 7)$ може приймати будь-яке значення із набору (1, 2, 3, 4, 5, 6), тобто є генератор Z_7 . Зворотною до отриманої таблиці є таблиця логарифмів (індексів):

Таблиця логарифмів (індексів)

n	0	1	2	3	4	5	6
ind_n	6	0	2	1	4	5	3

Користуючись таблицею індексів і експонент, неважко здійснювати деякі арифметичні операції.

$$2 \times 6 = 3^{ind 2} \times 3^{ind 6} = 3^{ind 2 + ind 6} = 3^{2+3} = 3^5 = 5,$$

$$3 \times 5 = 3^{ind 3} \times 3^{ind 5} = 3^{ind 3 + ind 5} = 3^{1+5} = 3^6 = 3^0 = 1,$$

$$5 \times 6 = 3^{ind 5} \times 3^{ind 6} = 3^{ind 5 + ind 6} = 3^{5+3} = 3^8 = 3^2 = 2.$$

В останніх двох виразах ми врахували, що $6 = 0(\bmod 6)$ і $8 = 2(\bmod 6)$. Іншими словами, множення у кільці чисел x і y здійснюється за таким правилом.

$$x \times y = \alpha^{ind x} \times \alpha^{ind y} = 3^{(ind x + ind y) \bmod p-1}.$$

Випишемо таблиці додавання і множення для Z_7 : Таблиці додавання і множення для Z_7

+	0	1	2	3	4	5	6	×	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6	0	0	0	0	0	0	0	0
1	1	2	3	4	5	6	0	1	0	1	2	3	4	5	6
2	2	3	4	5	6	1	2	2	0	2	4	6	1	3	5
3	3	4	5	6	0	1	2	3	0	3	6	2	5	1	4
4	4	6	0	1	2	3	4	4	0	4	1	5	2	6	1
5	5	6	0	1	2	3	4	5	0	5	3	1	6	4	2
6	6	0	1	2	3	4	5	6	0	6	5	4	3	2	1

Кільце Z_8 також можна перетворити у поле, але для цього необхідно визначити інші таблиці додавання і множення, попередньо побудувавши таблицю індексів. Такі поля (із власними таблицями додавання і множення) називаються *полями Галуа* $GF(p^k)$.

Для складання таблиці індексів поля $GF(2^3)$ скористаємося таким трюком. Візьмемо неприводимий поліном 3-го степеню із коефіцієнтами із Z_2 :

$$p(x) = x^3 + x + 1,$$

і розглянемо залишки від ділення степенів x^n на $p(x)$:

$$\left[\frac{x^0}{p(x)} \right] = 1 = 0001; \left[\frac{x^1}{p(x)} \right] = x = 0010; \left[\frac{x^2}{p(x)} \right] = x^2 = 0100; \left[\frac{x^3}{p(x)} \right] = x + 1 = 0011;$$

$$\left[\frac{x^4}{p(x)} \right] = x^3 + x = 0110; \left[\frac{x^5}{p(x)} \right] = x^2 + x + 1 = 0111;$$

$$\left[\frac{x^6}{p(x)} \right] = x^2 + 1 = 0101; \left[\frac{x^7}{p(x)} \right] = 1 = 0001.$$

Враховуючи, що у десятковому запису $001 = 1$, $010 = 2$, $011 = 3$, $100 = 4$, $101 = 5$, $110 = 6$, $111 = 7$, із цих залишків складемо таблицю індексів для $GF(2^3)$:

Таблиця індексів для $GF(2^3)$

n	0	1	2	3	4	5	6	7
2^n	1	2	4	3	6	7	5	1
$ind n$	7	0	1	3	2	6	4	5

Тепер, користуючись таблицею індексів, неважко побудувати таблицю множення для $GF(2^3)$:

Таблиця множення для $GF(2^3)$

$\oplus$	1	2	3	4	5	6	7	$\times$	1	2	3	4	5	6	7
1	0	3	2	5	4	7	6	1	1	2	3	4	5	6	7
2	3	0	1	6	7	4	5	2	2	4	6	3	1	7	5
3	2	1	0	7	6	5	4	3	3	6	5	7	4	1	2
4	5	6	7	0	1	2	3	4	4	3	7	6	2	5	1
5	4	7	6	1	0	3	2	5	5	1	4	2	7	3	6
6	7	4	5	2	3	0	1	6	6	7	1	5	3	2	4
7	6	5	4	3	2	1	0	7	7	5	2	1	6	4	3

Зазначимо, що таблиця додавання у даному випадку також будується інакше. В якості операції додавання використовується оператор побітового додавання XOR. -- " $\oplus$ ". Наприклад, $2 + 3 = 010 \oplus 011 = 001 = 1$, $5 + 6 = 101 \oplus 110 = 011 = 3$. Оскільки і за додаванням і за множенням наші елементи утворюють групи, дану множину елементів можна назвати полем $GF(2^3)$.

Аналогічним образом побудуємо поле Галуа $GF(2^4)$. Для цього візьмемо неприводимий поліном 4-го степеню із коефіцієнтами із Z_2 :

$$p(x) = x^4 + x + 1$$

і розглянемо залишки від ділення степенів x^n на $P(x)$:

$$\left[\frac{x^0}{p(x)} \right] = 1 = 0001; \quad \left[\frac{x^1}{p(x)} \right] = x = 0010; \quad \left[\frac{x^2}{p(x)} \right] = x^2 = 0100$$

$$\left[\frac{x^4}{p(x)} \right] = x + 1 = 0011; \quad \left[\frac{x^5}{p(x)} \right] = x^2 + x = 0110; \quad \left[\frac{x^6}{p(x)} \right] = x^2 + x^3 = 1100;$$

$$\left[\frac{x^8}{p(x)} \right] = x^2 + 1 = 0101; \quad \left[\frac{x^9}{p(x)} \right] = x^3 + x = 1010; \quad \left[\frac{x^{10}}{p(x)} \right] = 1 + x + x^2 = 0111.$$

Із урахуванням, що у десятковому запису $0001 = 1$, $0010 = 2$, $0011 = 3$, $1101 = 13$, $1110 = 14$, $1111 = 15$, із цих залишків складаємо таблицю індексів для $GF(2^4)$: Таблиця індексів для $GF(2^4)$

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2^n	1	2	4	8	3	6	12	11	5	10	7	14	15	13	9	1
indn	15	0	1	4	2	8	5	10	3	14	9	7	6	13	11	12

Тепер, користуючись таблицею індексів, неважко побудувати таблицю множення для $GF(2^4)$:

×	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	4	6	8	10	12	14	3	1	7	5	11	9	15	13
3	3	6	5	12	15	10	9	11	8	13	14	7	4	1	2
4	4	8	12	3	7	11	15	6	2	14	10	5	1	13	9
5	5	10	15	7	2	13	8	14	11	4	1	9	12	3	6
6	6	12	10	11	13	7	1	5	3	9	15	14	8	2	4
7	7	14	9	15	8	1	6	13	10	3	4	2	5	12	11
8	8	3	11	6	14	5	13	12	4	15	7	10	2	9	1
9	9	1	8	2	11	3	10	4	13	5	12	6	15	7	14
10	10	7	13	14	4	9	3	15	5	8	2	1	11	6	12
11	11	5	14	10	1	15	4	7	12	2	9	13	6	8	3
12	12	11	7	5	9	14	2	10	6	1	13	15	3	4	8
13	13	9	4	1	12	8	5	2	15	11	6	3	14	10	7
14	14	15	1	13	3	2	12	9	7	6	8	4	10	11	5
15	15	13	2	9	6	4	11	1	14	12	3	8	7	5	10

Табл.2.16 Таблиця множення для $GF(2^3)$

Додавання визначається як побітова сума за модулем 2 оператором XOR. У загальному випадку описана схема дозволяє для будь-якого p^k побудувати поле $GF(p^k)$.

2.3.2 Вихор Мерсенна

Вихор Мерсенна -- це генератор псевдовипадкових чисел, запропонований у 1997 японськими науковцями Макато Мацумото та Такудзі Нисимура. Вихор Мерсенна ґрунтується на властивостях простих чисел Мерсенна і забезпечує швидке генерування псевдовипадкових чисел.

Нехай x -- w -вимірні вектори над полем $F_2 = \{0,1\}$, що відповідають машинним словам розміром w . Вихор Мерсенна генерує послідовність векторів, які є псевдовипадковими цілими числами у діапазоні від 0 до $2^w - 1$. Шляхом ділення на $2^w - 1$ можна отримати псевдовипадкове число у діапазоні $[0,1]$.

Алгоритм ґрунтується на рекурентній формулі

$$x_{k+n} = x_{k+m} \oplus (x_k^u | x_{k+1}^l)A, \quad k = 0, 1, 2, \dots$$

n -- ціле, що позначає порядок рекурентності; m -- ціле, $0 \leq m \leq n$; A -- матриця розміром $w \times w$ із елементами із F_2 ; $|$ -- побітове АБО (XOR); $\oplus$ -- додавання за модулем 2 (XOR).

x_k^u позначає "старші $w-r$ біт" x_k , x_k^l "молодші r біт" x_{k+1} . Вектор $x_k^u | x_{k+1}^l$ є конкатенація старших $w-r$ біт x_k та молодших r біт x_{k+1} .

Візьмемо $(x_0, x_1, \dots, x_{n-1})$ в якості початкового заповнення. Генератор обчислює x_n при $k = 0$ тощо. Форма матриці A обрана з огляду на швидкість виконання множення на A :

$$A = \begin{bmatrix} 0 & 1 & & & & \\ 0 & 0 & 1 & & & \\ 0 & \dots & \dots & \ddots & & \\ \dots & & & & & 1 \\ a_{n-1} & a_{n-2} & \dots & \dots & \dots & a_0 \end{bmatrix}.$$

Обчислення xA зводиться до побітових операцій

$$xA = \begin{cases} x \gg 1 & x_0 = 0, \\ (x \gg 1) \oplus a & x_0 = 1, \end{cases}$$

де $x := x_k^u \mid x_{k+1}^l$, $k = 0, 1, \dots$; $a := (a_{w-1}, a_{w-2}, \dots, a_0)$; $x := (x_{w-1}, x_{w-2}, \dots, x_0)$.

Змінювання стану вихору Мерсенна здійснюється за лінійного співвідношення і визначається за допомогою лінійного перетворення B .

Відповідно до загальної теорії лінійних рекурентних послідовностей кожне значення в $n \times w - r$ матриці є лінійна рекурентна послідовність, що відповідає характеристичному поліному $\varphi_B(t)$ перетворення B . Матриця B має розміри $n \times (w - r) \times n \times (w - r)$:

$$B = \begin{bmatrix} 0 & I_w & \dots & 0 & 0 \\ \vdots & & & & \\ I_w & \vdots & & \vdots & \vdots \\ \vdots & & & & \\ 0 & \dots & \dots & I_w & 0 \\ 0 & 0 & \dots & 0 & I_w \\ S & 0 & \dots & 0 & 0 \end{bmatrix}; S = \begin{bmatrix} 0 & I_r \\ I_{w-r} & 0 \end{bmatrix}.$$

Для $l = 0, 1, \dots$ виконується

$$(x_{l+n}, x_{l+n-1}, \dots, x_{l+1}) = (x_{l+n-1}, x_{l+n-2}, \dots, x_l)B.$$

Характеристичний поліном $\varphi_B(t)$ має вигляд

$$\begin{aligned} \varphi_B(t) = & (t^n + t^m)^{w-r} \cdot (t^{n-1} + t^{m-1})^r + a_0(t^n + t^m)^{w-r} \cdot (t^{n-1} + t^{m-1})^{r-1} + \dots \\ & + a_{r-2}(t^n + t^m)^{w-r} \cdot (t^{n-1} + t^{m-1}) + a_{r-1}(t^n + t^m)^{w-r} + a_r(t^n + t^m)^{w-r-1} \\ & + \dots + a_{w-2}(t^n + t^m) + a_{w-1}. \end{aligned}$$

Послідовність досягає максимального періоду $2^{nw-r} - 1$ коли $\varphi_B(t)$ -- примітивний.

Метод ґрунтується на властивостях простих чисел Мерсенна. ``Вихор" -- перетворення, що забезпечує рівномірний розподіл ПВЧ (псевдовипадкові числа).

Числом Мерсенна називається натуральне число M_n :

$$M_n = 2^n - 1.$$

n -- просте число.

Приклад. 1, 3, 7, 15, 31, 63, 127, 255, 511, 1023, 2047, 4095, 8191, 16383, 32767.

Метод надає можливість генерувати послідовність двійкових псевдовипадкових цілих w -бітових чисел відповідно до рекурентної формули

$$X_{n+p} = X_{n+q} \oplus (X_n^r | X_{n+1}^l)A \quad n = 0, 1, 2, \dots,$$

де p, q, r -- цілі сталі; p -- степінь рекурентності, $1 \leq q \leq p$; X_n -- w -бітове двійкове ціле число;

$(X_n^r | X_{n+1}^l)$ -- двійкове ціле число, отримане конкатенацією чисел X_n^r і X_{n+1}^l , коли перші $(w-r)$ бітів взяті із X_n , а останні r бітів із X_{n+1} у тому самому порядку;

A -- матриця розміром $w \times w$, що містить 0 та 1, яка визначена через a ;

XA -- добуток, при обчисленні якого спочатку виконують операцію $X \gg 1$ (зсуву бітів на одну позицію вправо), якщо останній біт $X = 0$, а потім, коли останній біт $X = 1$, обчислюють $XA = (X \gg 1) \oplus a$,

$$a = (a_{w-1}, a_{w-2}, \dots, a_0),$$

$$X = (x_{w-1}, x_{w-2}, \dots, x_0),$$

$$A = \begin{bmatrix} 0 & 1 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \\ a_{w-1} & a_{w-2} & \dots & \dots & a_0 \end{bmatrix}.$$

Алгоритм Мерсенна складається з почергового виконання процедур *рекурсивної генерації* і *``закалювання''*. Рекурсивна генерація являє собою РЗДЗЗ (регістр зсуву зі зворотним лінійним зв'язком) із додатковою функцією для потоку вихідних бітів. Операція *``закалювання''* є процедура, що підсилює рівномірність розподілу на великих розмірностях бітових векторів.

Алгоритм.

1. Ініціюються значення u, h, a :

$$1a. u1 := (1, 0, \dots, 0) - (w-r) \text{ бітів}; \quad h1 := (0, 1, \dots, 1) - r \text{ бітів};$$

останній рядок матриці A :

$$a := (a_{w-1}, a_{w-2}, \dots, a_0).$$

1b. $x_0, x_1, \dots, x_{n-1}$ заповнюються початковими значеннями.

2. Обчислюється $(x_i^u | x_i^l)$

$$y := (x[i] \text{ AND } u1) \text{ OR } (x[(i+1)(\text{mod } n)] \text{ AND } h1).$$

3. Обчислюється нове значення X_i :

$$x[i] := \begin{cases} x_{(i+m)(\text{mod } n)} \text{ XOR } (y \gg 1) \text{ XOR } a, & y_0 = 1; \\ x_{(i+m)(\text{mod } n)} \text{ XOR } (y \gg 1) \text{ XOR } 0, & y_0 = 0. \end{cases}$$

4. Обчислюється $x[i]^T$.

$$y := x[i];$$

$$y := y \text{ XOR } (y \gg u),$$

$$y := y \text{ XOR } ((y \ll s) \text{ AND } b);$$

$$y := y \text{ XOR } ((y \ll t) \text{ AND } c);$$

$$Z := y \text{ XOR } (y \gg l).$$

Z подається на вихід як результат.

5. $i := (i+1)(\text{mod } n)$. Перехід до кроку 2.

Параметри алгоритму підібрані з метою досягнення найкращих властивостей. Параметри n, r підібрано так, що многочлен, що утворює, -- примітивний многочлен степеню 19937. w вибирається за розміром стандартного машинного слова -- 32 або 64 біта. Але для 65-бітової версії формула виглядає дещо інакше. Значення останнього рядку матриці A вибирається випадковим чином і подається на вхід алгоритму. Параметри закалювання підібрано так, щоб отримати гарний рівномірний розподіл.

$$n = 624, w = 32, r = 31, m(q) = 397;$$

$$a = 2567483615(99908B0DF_{16}), u = 11, s = 7, t = 15, l = 18,$$

$$b = 2636928640(9D2C5680_{16}), c = 4022730752(EFC60000_{16}).$$

Даний метод не призначений для отримання криптографічно стійких послідовностей випадкових чисел.

2.4 Створення випадкового числа з біоелектричних та фізичних сигналів

Детерміновані генератори випадкових чисел генерують числа швидко, легко, недорого та є апаратно незалежні рішення. Статистичні якості цих чисел близькі до ідеальних.

Випадкові числа широко використовуються в таких сферах, як криптографія та передача даних, ігри на удачу, безпечне спілкування, моделювання та програмування ігор, де важлива генерація ключів. Генератори випадкових чисел можна розділити на два класи: TRNG (Генератор справжніх випадкових чисел) та PRNG (генератор псевдовипадкових чисел). Випадкові числа, створені програмним забезпеченням, можна визначити за певною математичною моделлю. З іншого боку, можна генерувати числа за допомогою апаратного забезпечення джерелом шуму, поведінку якого неможливо передбачити.

Вимоги до випадкових чисел.

R_1 -- RNG повинні генерувати випадкові числа, що мають хороші статистичні властивості на виході, для використання в якому криптографічні програми.

R_2 -- Якщо зломисник знає підгенератори випадкових чисел, не можна допускати обчислення або прогнозувати передумови та послідовні випадкові числа з високою точністю.

R_3 -- Неможливо передбачити або обчислити раніше згенеровані випадкові числа з високою точністю за допомогою враховування відомого поточного значення внутрішнього стану PNG або не вимагаючи інформації про його внутрішній стан.

R_4 -- Неможливо передбачити або обчислити наступні випадкові числа з високою точністю, враховуючи відоме поточне значення внутрішнього стану PNG або без вимагання інформації про його внутрішній стан.

TRNG (True Random Number Generators, справжні генератори випадкових чисел), які є генераторами невизначених випадкових чисел, повільніші, дорожчі та апаратно-залежні рішення порівняно з PRNG. На відміну від PRNG, немає необхідності включати додаткові компоненти до конструкцій системи TRNG для вимог R_2, R_3 та R_4 . Через непередбачуваність випадкових чисел, утворених у результаті використання високого рівня шуму джерел з високою ентропією в TRNG, передбачається, що R_2 вимога виконана. Якщо вимога R_2 задоволена, то це передбачає, що вимоги R_3 та R_4 також задоволені.

Для того, щоб задовольнити вимогу R_1 у TRNG, проводиться післяобробка. застосовуються методи до випадкових чисел, отриманих за допомогою відбору проб з джерел шуму. Це виключає статистично слабкі місця випадкових чисел на виході TRNG.

Метод кодування для генерації випадкових чисел за допомогою сигналів ЕЕГ -- генератор чисел, який можна використовувати за низькою вартістю та реальних додатків, заснований на TRNG.

2.4.1 Методи генерації випадкових чисел

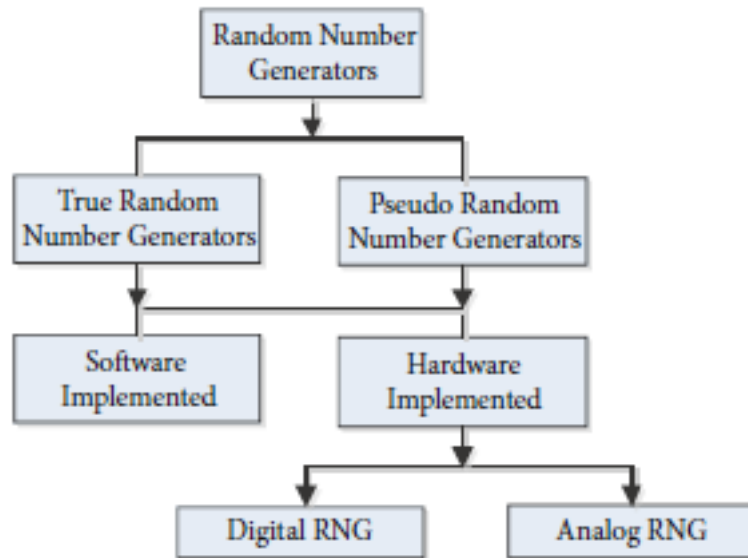


Рис.2.17 Класифікація генераторів випадкових чисел

PRNG

Загальна архітектура проектування PRNG показана на рис. 2.17. $r_1, r_2, \dots, r_n \in R$ -- генератор випадкових чисел, $S_n \in S$ -- внутрішні стани чистого PRNG, P_s -- розподіл ймовірностей випадкових seed. PRNG генерує випадкове число r_n з поточного стану S_n , $\Psi: S \rightarrow R$, $r_n = \Psi(S_n)$. Далі, із використанням функції переходу Φ стан S_n оновлюється, оскільки $S_{n+1} = \Phi(S_n)$. S_0 є перший внутрішній стан і генерується $S_1 = \Phi(S_0)$.

Генератори випадкових чисел з хорошою статистикою генеруються розширенням цих параметрів за допомогою детермінованих способів [19].

TRNG

Загальна архітектура проектування TRNG показана на рис. 2.18.

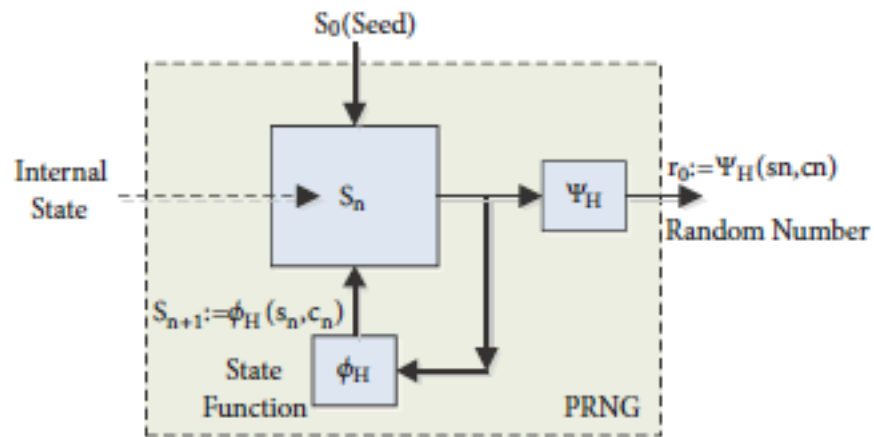


Рис.2.18 Загальна архітектура проектування PRNG

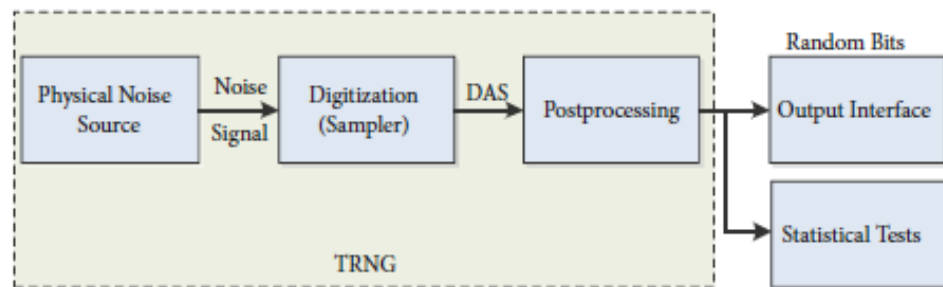


Рис.2.19 Структура блоку TRNG

Значення, що обчислені із вибірки джерел шуму, називаються оцифрованими аналоговими сигналами (DAS). DAS випадкові числа відповідають окремому випадку чисто випадкових чисел, і вони піддаються алгоритмізації післяопрацювання, щоб зменшити їх потенційні слабкості. Але швидкість виведення бітів зменшується і робоча швидкість зменшується.

2.4.2 Біоелектричні та фізичні сигнали

Біоелектричні сигнали - це низькі амплітудні шуми між 100 і 1 мВ і виводяться з тіла за допомогою електродів. Спектри частот таких сигналів знаходяться в низькочастотних діапазонах 0,1 Гц ~ 2000 Гц. Амплітудні та частотні характеристики різних біоелектричних сигналів, взяті з тіла. ЕЕГ-сигнал: амплітуда -- 2–100 μV ; частота - - 0,5 -- 50 Гц.

Під час мозкової діяльності виробляються неперервні ритмічні електричні потенціали, а також генеруються електричні сигнали внаслідок активності рецептора. Амплітуди хвиль ЕЕГ -- у діапазоні від 5 до 400 В і їх частоти змінюються між 0,5 і 100 Гц.

Електроміографія (ЕМГ) - це метод неврологічного обстеження, заснований на дослідженні електричних потенціалів нервів і м'язів. ЕМГ проводиться двома способами за допомогою поверхневих електродів та голчастих електродів. У тестах використовують поверхневі електроди, електроди прикріплені до поверхні шкіри.

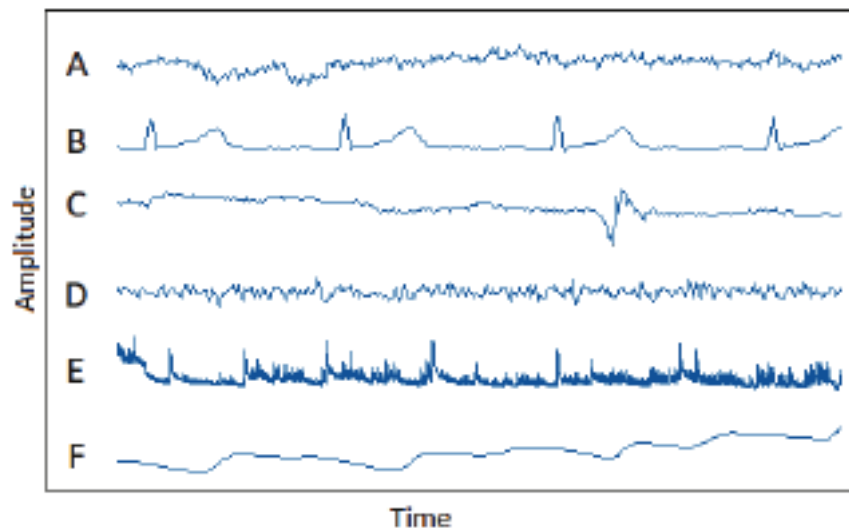


Рис.2.20 Біоелектричні та фізичні сигнали з бази даних

Поверхнева ЕМГ може допомогти контролювати м'язові та нервові розлади. У випробуваннях з використанням голчастих електродів проколюється в м'язову тканину і записуються електричні сигнали на м'язових волокнах. Амплітуди хвиль ЕМГ змінюються від 100 В до 1 мВ, а їх частоти коливаються від 10 до 500 Гц.

Електроокулограммічні сигнали (ЕОГ)-- це сигнали рогівки та сітківки ока між рогівкою та сітківкою ока, утворені рухами очей і викликані гіперполяризацією та деполяризацією. ЕОГ сигнали приймаються за допомогою розміщених навколо електродів очі. Сигнали ЕОГ знаходяться в смузі частот близько 0,1–10 Гц, а їх амплітуди становлять близько 0,001–100 мВ. Горизонтальні та вертикальні сигнали ЕОГ змінюються залежно від руху очей. Один ступінь руху викликає зміну горизонталі на 16 В амплітуди і 14 В у вертикальній амплітуді.

Із фізичних сигналів використовується об'ємний пульс крові (BVP) для вимірювання частоти серцевих скорочень. Отримано вимірювання BVP за допомогою датчика фотоплетізмографії (PPG). Це датчик вимірює зміни об'єму крові відповідно до змін в частоті серцевих скорочень в артеріях і капілярах і кровотечі.

Сигнал GSR є одним з найбільш чутливих показників емоційної стимуляції, щоб показати, чи є люди під стресом. Він дає інформацію про провідність шкіри. Інший фізичний сигнал, дихання, викликається різниця між вдихаючим повітрям і видихуванім повітрям.

На рис. 2.20 показані біоелектричні та фізичні зразки сигналів в базі даних BNCIHORIZON2020. У цій фігурі, шість рядів сигналів зверху вниз - це зразки ЕЕГ, ЕОГ, ЕМГ, GSR та сигнали дихання артеріального тиску з наборів даних A, B, C, D, E, F відповідно.

2.4.3 Генерація особистих ідентифікаторів з біоелектричних та фізичних сигналів.

У цьому дослідженні для створення випадкових чисел на основі TRNG використовувалися біоелектричні та фізичні сигнали, отримані з бази даних BNCIHORIZON2020. Загальна кількість бази даних 15. Ці дані -- чотири ЕЕГ, чотири ЕМГ, чотири ЕОГ, один об'ємний пульс крові, один GSR (відповідь гальванічної шкіри) і одне дихання. Для даних ЕЕГ сигнали в базу даних записували за допомогою тридцяти

двох активних Ag/AgCl електродів з частотою дискретизації 512 Гц відповідно до розширеної міжнародної системи 10–20. GSR, дані про об'єм пульсу крові та дані про дихання, записувалися одночасно з біоелектричними сигналами.

Створення реальних випадкових чисел на основі цих даних включає в себе три кроки, як показано на рис. 2.21. Цими кроками є нормалізація, оцифрування (семплер) та постпроцесінг.

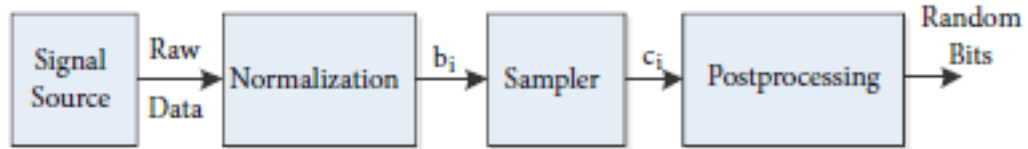


Рис.2.21 Кроки генерування чисел

При нормалізації всі дані, отримані з джерел шуму (вихідні сигнали) спершу слід перетворити на двійкове число системи. Для цього операції, дані з їх допомогою застосовуються математичні пояснення нижче. Нехай кожен сигнал отриманий від окремих осіб $x = (x_1, x_2, \dots, x_n)^T$. Кожен зразок виражається як 5 біт за допомогою модульної арифметики

$$y_i = x_i \pmod{32}, \quad i = \overline{1, n}. \quad (29)$$

Для виробництва 0 і 1 з 5-розрядних генераторів y_i генератор z_i отримано згідно (30). Кожен елемент генератора z_i знаходиться всередині діапазону [0--5].

$$z_i = \sum_{k=1}^5 y_{i_k}. \quad (30)$$

Нарешті, b_i бінарний генератор отримується за допомогою z_i за допомогою (31).

$$b_i = z_i \pmod{2}, \quad i = \overline{1, n}. \quad (31)$$

Після цього кроку вибірка застосовується до отриманого b_i -генератора від біоелектричних та фізичних сигналів, які використовуються як джерело шуму. У цій статті логістика для вибірки була використана карта, що представляє неперіодичну поведінку. Логістична карта, що представляє хаотичну поведінку, визначається як

$$a_{i+1} r \star a_i \star (1 - a_i), \quad i = 0, 1, 2, \dots, \quad (32)$$

де r -- системний параметр, a_1 - початкове або початкове значення, а i -- кількість ітерацій. Числа, що виробляються логістичною картою та в діапазоні [0, 1] є дійсні. Нехай ці числа -- $a1, a2, a3$.

Отримані біти сигналів

b_i -- (Нормовані сигнали)

0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0

c_i (Вибірковий сигнал, отриманий з логістичної карти)

0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1

d_j -- (дискретизований сигнал)

1, 1, 1, 0, 1 1, 0 0, 1, 0

Вираз " $a_i \geq 0,5$ ", вироблений в i -й ітерації, вказує, що вихід логістичної карти

передбачається рівним 1; в іншому випадку це 0. Для $r = 3,91$ і $a_1 = 0,2$, перші 15 значень: 0,2, 0,625, 0,916, 0,301, 0,823, 0,568, 0,959, 0,153, 0,508, 0,977, 0,087, 0,311, 0,837, 0,533 та 0,973. Вироблений сигнал дискретизації буде таким: 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1.

Коли сигнал дискретизації стає 1, тоді $d_j = b_i$ задовольняється. Коли він дорівнює 0, сигнал b_i не враховується.

Доопрацювання застосовується для поліпшення статистичних властивостей вибірових генераторів випадкових чисел. Структури функцій XOR, фон Неймана, LFSR і Nash зазвичай використовуються для післяобробки [21, 28]. У цьому дослідженні використаний постпроцесор XOR. Коли послідовні числа 0,1 і 1,0 з XOR у генераторі випадкових чисел, дійсне випадкове число вважається 1, а інакше - 0. Коли після опрацювання XOR застосовується до вибірки 1110110010 генератор чисел буде 01001.

2.4.4 Завдання для самостійної роботи

Задача 1. Визначте закон розподілу випадкової величини, якщо ряд значень випадкової величини, що спостерігається, такий:

10, 7, 8, 8, 8, 8, 10, 10, 9, 10, 7, 10, 7, 10, 9, 9, 8, 8, 8, 9, 6, 6, 5, 8, 7, 8, 8, 7, 8, 9, 9, 9, 8, 7, 8, 7, 6, 9, 9, 8.

Задача 2. Визначте закон розподілу випадкової величини, якщо ряд значень випадкової величини, що спостерігається, такий:

121, 76, 65, 252, 85, 36, 137, 80, 40, 48, 99, 101, 113, 124, 111, 66, 19, 170, 148, 42, 107, 184, 83, 0, 29, 34, 74, 103, 70, 83, 127, 111, 120, 42, 45, 45, 80, 34, 0, 163, 139, 107, 49, 53, 56, 113, 80, 165, 31, 118.

Задача 3. Побудуйте апроксимуючу функцію виду $y = b_0 + b_1 \frac{1}{x} + b_2 \frac{1}{x^2}$ за такими результатами спостережень:

X	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9
Y	12	13	11	8	5	9	2	7	4	5	5	6	7	9	5	5	4

Задача 4. Побудуйте апроксимуючу функцію виду

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

за такими результатами спостережень:

X_1	48	2	4	22	31	22	32	2	21	17	43	11	24	23	22
X_2	18	14	13	17	16	4	17	17	18	8	17	9	7	11	15
X_3	25	8	20	11	16	16	22	8	24	15	24	13	19	16	11
Y	381	8	31	167	242	182	252	5	164	138	341	88	197	183	169

Порядок виконання роботи

а) Отримати явну формулу для моделювання випадкової величини із законом розподілу, що задано в табл. (100). При необхідності знайти невідомі параметри.

б) Моделювання випадкової величини $X = \{x_i\} = [x_1, x_2, \dots, x_n]$ виконати за допомогою послідовності чисел $R = \{r_i\} = [r_1, r_2, \dots, r_n]$, що отримано указаним в таблиці (100) способом побудови випадкових чисел.

в) Обчислити критерій χ^2 при $k = 16$ або $k = 21$ для послідовних випадкових

чисел $X = [x_1, x_2, \dots, x_n]$ і зробити висновок про відповідність величини, що змодельована, даному закону розподілу.

г) Вивести на екран таблицю, що заповнена інтервальним статистичним рядом, гістограму, що побудована по табл. (100), теоретичний і практичний графіки випадкових величин, що моделюються за заданим законом розподілу, значення критерію χ^2 .

Варіанти завдань

Варіант	Закон розподілу	Спосіб побудови
1	Експоненційний, $\lambda = 2$	Метод зворотних функцій,
2	Нормальний, $N(0;1)$	Метод відбору
3	Розподіл Пуассона; $\lambda = 9$	
4	Нормальний, $(0; 0.81)$	Метод відбору
5	Нормальний, $N(0;1)$	Апроксимація розподілу
6	Γ -розподіл; $\lambda = 3$	
7	Логарифмічно-нормальний,	
	$\sigma_Y = 1, m_Y = 1$	Метод відбору
8	β -розподіл; $m = 2, p = 1.5$	
9	Нормальний, $N(0;1)$	Апроксимація
10	Дискретний розподіл $k = 10;$	
	$p = [0,5; 0,8; 0,7; 0,2; 0,04;$	
	$0,09; 0,3; 0,95; 0,97; 0,35]$	
11	Нормальний розподіл,	Процедура 4
12	Нормальний, $N(0;1)$	Методи суперп. і відбору,
13	Експоненційний, $\lambda = 0,4$	Метод зворотних функцій,
14	Нормальний; $N(0;1)$	Процедура Марсальї і Брея,
19	Γ -розподіл; $\lambda = 4$	

2.5 Контрольні питання

1. У чому полягає сутність методу статистичного моделювання систем на ЕОМ?
2. Яка послідовність випадкових чисел використовується у якості базової при статистичному моделюванні на ЕОМ?
3. Чому послідовності чисел, що генеруються на ЕОМ, називаються псевдовипадковими?

4. Що собою являють конгруентні процедури генерування послідовностей?
5. Які існують способи генерування послідовностей випадкових чисел із заданим законом розподілу?
6. У який спосіб дослідник може визначити закон розподілу випадкової величини?
7. Дайте формулювання задачі ідентифікації закону розподілу.
7. Які критерії застосовуються для визначення відповідності обраного закону розподілу випадковим числам, що досліджуються?
8. Як розраховується критерій χ^2 , λ -критерій?
9. За яких умов для визначення відповідності обраного закону розподілу досліджуваним випадковим числам застосовується критерій χ^2 ?
10. Опишіть дії, що виконуються дослідником при визначенні закону розподілу випадкової величини.
11. Дайте формулювання задачі ідентифікації функціональної залежності.
13. За якої умови задача апроксимації не має розв'язку?
14. За якими міркуваннями дослідник обирає вид апроксимуючої функції?
13. Чи існують обмеження на вид функціональної залежності, що дослідник висуває в задачі апроксимації? Які?
14. За яким критерієм визначаються параметри функціональної залежності?
15. Який зміст критерію найменших квадратів?
- 16) За якими формулами виконується кореляційно-регресійний аналіз?
- 17) Який показник характеризує щільність зв'язку?
- 18) Що оцінюють за допомогою шкали Чеддока?
- 19) За яким критерієм оцінюється значимість параметрів функціональної залежності?
- 20) За яким критерієм визначається істотність індексу кореляції?
16. За якими формулами визначається довірчий інтервал параметра функціональної залежності?
17. Яку інформацію надає знання довірчого інтервалу?
17. В яких випадках використовується апроксимація за критерієм χ^2 ?

Розділ 3 Теорія масового обслуговування

3.1 Загальні засади теорії

Під системою масового обслуговування (СМО) мають на увазі динамічну систему, що призначена для ефективного обслуговування потоку заявок (вимог до обслуговування) за обмежень на ресурси системи. Моделі СМО зручні для опису сучасних обчислювальних систем, наприклад, процесор -- основна пам'ять, канал вводу -- виводу тощо.

Сукупність взаємопов'язаних СМО називається мережею масового обслуговування. Основні задачі, що вирішуються у рамках теорії масового обслуговування:

- аналіз, тобто визначення кількісних характеристик СМО;
- синтез оптимальної структури СМО при заданих характеристиках та обмеженнях на параметри елементів структури.

На вхід СМО надходять заявки на обслуговування -- вхідний потік. У розімкнутих СМО число заявок вважається необмеженим. Для замкнутих СМО характерна скінченна кількість заявок. Заявки, що обслужені, повертаються у джерело і через деякий час можуть знову з'явитися на вході СМО. Поведінка джерела у замкнутих СМО є певна функція стану СМО.

Крім апаратних засобів (процесори, пам'ять, пристрої переривання, периферійні пристрої) до складу СМО входять програмні засоби, що містять прикладні і системні програми.

Прикладні програми, що управляють, реалізують алгоритми обробки інформації, їхнє виконання процесорами розглядається як обслуговування заявок, що надходять у ВС (диспетчерування) і виконуються, як правило, одним процесором. Зазвичай відокремлюють дві основні системні програми, що управляють: Диспетчер №1 (Д1) і Диспетчер №2 (Д2), що реалізують, відповідно, дисципліни очікування і обслуговування.

Система функціонує так. Поява на вході СМО заявки ініціює виконання програми Д1, яка розпізнає пріоритет заявки, що надійшла, і ставить її у відповідну чергу O_i , що реалізована у спеціально зарезервованій області пам'яті. Д2 аналізує стан черг $O_1, O_2, \dots, O_N$ і стан каналів $K_1, K_2, \dots, K_m$ (процесорів). За наявності вільного каналу Д2 вибирає заявку, що має переважне право на обслуговування, та ініціює відповідний канал і потрібну прикладну програму. Ініціювання здійснюється у моменти завершення виконання прикладних програм і програми Д1.

Процес надходження заявок у СМО є у загальному випадку випадковий і може розглядатися як потік однорідних подій, що трапляються через випадкові проміжки часу. Випадкові інтервали часу між надходженням заявок можуть підпорядковуватися різним законам розподілу. Найрозповсюдженішим у теорії масового обслуговування є найпростіший потік, тобто потік, в якому інтервал часу між двома сусідніми заявками підпорядковується експоненціальному закону розподілу із інтенсивністю λ :

$$f(\tau) = \lambda e^{-\lambda\tau}.$$

Припущення про найпростіший потік надає можливість отримати аналітичні залежності характеристик СМО від параметрів потоку, що входить.

Найпростіший потік у теорії масового обслуговування відіграє таку саму роль, що й нормальний закон розподілу випадкових величин у теорії ймовірностей: при додаванні кількох незалежних, ординарних, стаціонарних випадкових потоків утворюється сумарний потік, що наближається за своїми властивостями до найпростішого.

Якщо потік, що входить, є сукупність M потоків заявок різних типів із інтенсивностями λ_i ($i = \overline{1, M}$), його можна характеризувати сумарною інтенсивністю

$$\lambda = \sum_{i=1}^M \lambda_i.$$

За ступенем важливості заявки поділяють на класи із пріоритетами K_i ($i = \overline{1, N}$), причому найвищий пріоритет мають заявки першого класу.

Параметри структури СМО.

Кожна СМО має визначену структуру, що характеризується сукупністю параметрів. Основний компонент структури СМО є канали обслуговування. Залежно від числа каналів розрізняють одноканальні та многоканальні СМО. Продуктивність каналу обслуговування є зворотна до терміну обслуговування заявки, що дорівнює часу, який необхідний каналу для обслуговування заявки. У загальному випадку це випадкова величина із *функцією розподілу* $F(\tau_{zv})$, *густиною розподілу* $f(\tau_{zv})$ і *математичним сподіванням* $\overline{\tau_{zv}}$.

Типи заявок розрізняються або законом розподілу, або тільки математичними сподіваннями за однакових законів розподілу. При цьому приймається припущення про незалежність тривалості обслуговування різних заявок одного типу.

Разом із математичним сподіванням тривалості обслуговування використовується поняття *інтенсивності потоку обслуговування* $\mu = 1/\tau_{zv}$, що характеризує кількість заявок, які можна обслужити за одиницю часу каналом обслуговування, що постійно завантажений. При моделюванні СМО найчастіше використовують тривалість обслуговування із *експоненційною густиною розподілу*:

$$f(\tau_{zv}) = \mu e^{-\mu \tau_{zv}}.$$

Інший важливий компонент структури СМО є черга, параметром якої є число місць n у черзі. У пріоритетних системах загальна черга може бути поділена на кілька черг за числом пріоритетів, що розрізняє система, для кожної із яких має бути указане число місць n_i ($i = \overline{1, N}$). Залежно від числа місць у черзі розрізняють СМО із відмовами та СМО без відмов. У СМО із відмовами число місць у черзі скінченне і внаслідок ймовірного характеру як потоку, що входить, так і процесів обслуговування, існує ймовірність того, що для заявки, яка надійшла на вхід СМО, виявиться, що усі канали зайняті обслуговуванням і усі місця у черзі зайняті заявками, що чекають у черзі, тобто вона отримає відмову. У СМО без відмов заявка або одразу приймається на обслуговування, коли у момент її надходження вільний бодай один канал, або безумовно приймається у чергу на обслуговування.

Параметри потоку, що виходить.

Цей потік у загальному випадку розпадається на потік заявок, що обслужені, і потік, заявок, які втрачені, кожний з яких характеризується законом розподілу тривалості інтервалу між сусідніми заявками.

Якщо потік, що виходить, містить заявки M типів з інтенсивностями потоку заявок типу i , ($i = \overline{1, M}$), потік, що виходить, можна характеризувати сумарною інтенсивністю потоку заявок, що обслужені

$$\lambda^o = \sum_{i=1}^M \lambda_i^o,$$

та сумарною інтенсивністю потоку заявок, що втрачено

$$\lambda^L = \sum_{i=1}^M \lambda_i^L.$$

$$\lambda_i^O + \lambda_i^L = \lambda_i.$$

Своєю чергою, потік втрачених заявок може складатися із потоку заявок, що отримали відмову, і потоку "нетерплячих" заявок, що залишили систему, оскільки час їхнього перебування перевищило допустиму величину, тобто

$$\lambda_i^P = \lambda_i^A + \lambda_i^O.$$

Формально система масового обслуговування (СМО) -- це система, що складається із одного або кількох джерел запитів (заявок, вимог) на виконання певних дій, кількох приладів обслуговування (каналів обслуговування), що виконують ці дії відповідно з певними правилами (дисципліни обслуговування) за запитами, що надходять у систему.

Джерело запитів визначається як зовнішня щодо СМО система, із якої запити надходять до СМО для обслуговування.

Вхідний потік. Запити надходять у канал обслуговування з джерела у моменти часу $t_0 < t_1 < \dots < t_k < \dots$. Інтервали часу $\xi_i = t_i - t_{i-1}$, $i > 1$ між послідовними моментами надходження запитів є випадкові величини. ξ_i утворюють послідовність незалежних і однаково розподілених випадкових величин з функцією розподілу $F(t) = P\{\xi_i < t\}$.

Процес обслуговування. Нехай η_k -- тривалість обслуговування k -го запиту. η_k ($k = 1, 2, \dots$ -- незалежні однаково розподілені випадкові величини з функцією розподілу $B(t) = P\{\eta_k < t\}$. $B(t)$ -- розподіл тривалості обслуговування. Найпоширеніший закон розподілу є експоненційний з функцією розподілу $B(t) = 1 - e^{-\mu t}$. Параметр μ -- інтенсивність обслуговування. Середній час обслуговування дорівнює $1/\mu$.

Канали обслуговування. СМО може мати один або більше каналів (ліній, приладів) обслуговування. Прилади обслуговування можуть бути однорідними і неоднорідними. У СМО із однорідними приладами усі прилади обслуговують запити однаково. У СМО із неоднорідними приладами прилади відрізняються один від іншого деякими параметрами, наприклад, інтенсивністю обслуговування.

Дисципліна обслуговування. Правило, згідно з яким запити вибираються для обслуговування, складають дисципліну обслуговування. Усі дисципліни поділяються за наданих переваг на безпріоритетні і пріоритетні. Безпріоритетні дисципліни обслуговування поділяються на дисципліну обслуговування у порядку надходження, у зворотному порядку, із випадковим вибором із черги та циклічну дисципліну обслуговування. Дисципліна обслуговування у порядку надходження (FIFO -- First Input First Output) є найпоширеніша. Дисципліна обслуговування у зворотному порядку (інверсному) (LIFO -- Last Input First Output) використовуються в операційних системах під час обробки переривань та організації стеків. Циклічні дисципліни обслуговування використовуються під час обробки інформації у режимі реального часу.

За наявності певної ознаки СМО можна класифікувати так, [26]:

1. За кількістю вимог, що надходять за одиницю часу, на системи з *ординарними і неординарними потоками вимог*.

2. За зв'язком між вимогами -- на системи *без післядії і з післядією*. Якщо ймовірність надходження вимог у систему не залежить від кількості вимог, що уже надійшли, тобто не пов'язана із передісторією процесу, що досліджується, отримаємо систему без післядії. Прикладом системи із післядією може слугувати потік студентів, що здають залік викладачу.

3. За реакцією вимоги на зайнятість каналів -- на системи з відмовами й очікуванням.

4. Системи з очікуванням розподіляються на системи з обмеженим і необмеженим очікуванням.

5. За часом обслуговування вимоги -- на системи з детермінованим і випадковим часом обслуговування.

6. За кількістю каналів обслуговування -- на одно- і багатоканальні системи.

Випадковий процес, що протікає в СМО, полягає у тому, що система у випадкові моменти часу переходить із одного стану в інший: змінюється число зайнятих каналів, число заявок, що перебувають у черзі, тощо.

Процес називається *процесом із дискретними станами*, якщо його можливі стани $S_1, S_2, \dots, S_n$ можна заздалегідь перелічити, а перехід системи із стану у стан здійснюється миттєво (стрибкоподібно).

Процес називається *процесом із неперервним часом*, якщо моменти можливих переходів системи із стану у стан випадкові (не фіксовані заздалегідь).

Процес роботи СМО є випадковий процес із дискретними станами і неперервним часом.

Показники ефективності СМО.

Тут маються на увазі кількісні показники, що частково характеризують рівень виконання СМО функцій, що покладено на неї. На ґрунті цих показників можна побудувати деякий критерій ефективності, що характеризує ефективність СМО за обмежень на її параметри. Розглянемо найвживаніші з них. Зауважимо, що усі показники відображають можливості СМО по обслуговуванню заявок, але не характеризують якість самого обслуговування.

Ймовірність обслуговування

$$P_{\text{іа}} = \lambda^0 / \lambda.$$

Іноді ймовірність обслуговування називають відносною пропускнуою спроможністю.

Ймовірність втрати $P_{\text{А}}$ характеризує ймовірність того, що довільно обрана заявка із потоку, що входить з інтенсивністю λ , виявиться у потоці заявок, що втрачено з інтенсивністю $\lambda_{\text{А}}$:

$$P_{\text{А}} = \lambda^{\text{А}} / \lambda = (\lambda - \lambda^0) / \lambda = 1 - P_{\text{іа}}$$

Середній час очікування $t_{\text{і+}}$ заявки (середній час перебування заявки у черзі) є математичне сподівання часу очікування. Середній час очікування у загальному випадку є сума двох складових:

- $\overline{t_{\text{і+}}^{\text{І}}}$ -- середнього початкового часу очікування, що дорівнює проміжку часу між моментом появи заявки на вході СМО і моментом першого призначення заявки на обслуговування або уходу із черги;

- $\overline{t_{\text{і+}}^{\text{ІІ}}}$ -- середнього часу очікування у перерваному стані, що дорівнює сумі проміжків часу між моментами надходження заявки, обслуговування якої було перерване, у чергу та моментами або повторного призначення заявки на обслуговування (продовження обслуговування заявки з того стану, в якому вона перебувала на момент чергового переривання), або втрати заявки за рахунок уходу її:

$$\overline{t_{\text{і+}}} = \overline{t_{\text{і+}}^{\text{І}}} + \overline{t_{\text{і+}}^{\text{ІІ}}}.$$

Середній час перебування заявки в СМО $t_{\text{Н}}$ заявки є математичне сподівання часу

перебування заявки в СМО. t_N дорівнює проміжку часу від моменту надходження заявки на вхід СМО до моменту появи її на виході і пов'язане із тривалістю процесів очікування $t_{\bar{t}}$ та обслуговування $t_{\bar{a}}$.

Середня довжина черги $\bar{I}$. Для визначення $\bar{I}$ потрібно знати сукупність ймовірностей $P_{i\bar{i}}$, ($i = \overline{1, n}$), тобто ймовірностей перебування у черзі i заявок.

$$\bar{I} = \bar{t}_{\bar{t}} \lambda.$$

Середнє число зайнятих каналів обслуговування $\bar{K}$ дорівнює математичному сподіванню числа зайнятих обслуговуванням каналів і характеризує ступінь завантаженості системи обслуговування. Для визначення $\bar{K}$ треба знати завантаження P_{SI} -- ймовірність того, що у довільний момент часу обслуговуванням будуть зайняті усі m каналів:

$$P_{SI} = \bar{K}/m.$$

Середнє число заявок у системі $\bar{Z}$ є математичне сподівання числа заявок, що одночасно перебувають у черзі або у каналі обслуговування.

$$\bar{Z} = \bar{I} + \bar{K}.$$

Для СМО без втрат маємо

$$\bar{Z} = \bar{I} + \bar{K} = \lambda(\bar{t}_{\bar{t}} + \bar{t}_{\bar{a}}) = \bar{t}_C.$$

Функціонування широкого класу систем можна подати як процес переходу із одного стану в інший під впливом деяких збурень. Наприклад, функціонування ЕОМ характеризується тим, що у кожний момент часу обробкою інформації зайняті певні блоки. Цей процес можна розглядати як процес переходу системи із одного стану в інший.

Нехай є деяка фізична система A , що у процесі функціонування може приймати різні стани A_i . Якщо стани системи змінюються у часі випадковим чином, процес змінювання станів можна розглядати як випадковий процес, що описується випадковою функцією $X(b)$. Повна множина станів системи може бути або скінченною ($i = \overline{1, n}$), або нескінченно великою.

Завданням дослідження СМО є встановлення залежностей базових характеристик: середній час очікування, інтенсивностей обслуговування на приладах, дисципліни обслуговування.

Розрізняють кілька підходів до дослідження СМО: диференціальний підхід, метод фаз, метод вкладених ланцюгів Маркова, метод додавання змінних.

3.2 Методи дослідження СМО

3.2.1 Диференціальний метод

Диференціальний метод застосовують для дослідження СМО з пуассонівським потоком подій, тобто час обслуговування й інтервали часу між надходженням запитів підпорядковані експоненційному розподілу.

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

де λ -- параметр потоку.

У цьому разі процес змінювання кількості запитів є *марківський*.

Необмежена послідовність дослідів із єдиною можливістю і попарно несумісними

витоками $(s_1, s_2, \dots, s_n)$ зветься ланцюгом Маркова, якщо ймовірність будь-якого із цих витоку в черговому досліді однозначно визначається результатом безпосередньо попереднього досліді.

Сукупність несумісних витоку $s = (s_1, s_2, \dots, s_n)$ зветься вектором стану системи, компоненти якого утворюють повну групу подій: $s_1 + s_2 + \dots + s_n = 1$.

Позначення: $p_j^{(0)}$ -- ймовірність i -го витоку у першому досліді ($j = \overline{1, m}$).

$$p_1^{(0)} + p_2^{(0)} + \dots + p_m^{(0)} = 1.$$

$p_{ji}^{(0)}$ -- ймовірність i -го витоку в n -му досліді за умови, що у $(n-1)$ -му досліді настав i -й витоку, $n = 2, 3, \dots$; $i, j = \overline{1, m}$; $p_{1i}^{(n)} + p_{2i}^{(n)} + \dots + p_{mi}^{(n)} = 1$. Ймовірностями, що запроваджено, сповна описується ланцюг Маркова.

Ланцюг Маркова зветься однорідним, якщо ймовірності $p_{ij}^{(n)}$ не залежать від n , у цьому разі вони позначаються без верхнього індекса p_{ij} . Числа p_{ij} зветься ймовірностями переходів, а

$$P = [p_{ij}] = \begin{vmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{vmatrix}$$

-- матрицею переходу.

По матриці переходу можна побудувати граф станів, якщо в якості вузлів взяти стани $s = (s_1, s_2, \dots, s_n)$ системи, а стрілками -- можливі переходи із одного стану в інший.

Наприклад, для матриці переходу

$$P = \begin{vmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 \\ p_{21} & 0 & 0 & 0 & 0 \\ p_{31} & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{43} & 0 & 0 \\ 0 & 0 & p_{53} & 0 & p_{55} \end{vmatrix}$$

граф стану має вигляд  Граф стану

Позначимо: $p_{ij}(n)$ -- ймовірність того, що через n кроків здійсниться перехід від s_j до s_i ($i, j = \overline{1, m}$), $P(n)$ -- матрицю переходів за n кроків, тобто матрицю, яка складається із $p_{ij}(n)$. Якщо у початковий момент система мала стан

$s^0 = (s_1, s_2, \dots, s_n)$, на наступному кроці вона перейде у стан

$$s^1 = P s^0.$$

На другому кроці стан системи є:

$$s^2 = P s^1 = P \cdot P s^0 = P^2 s^0.$$

Вочевидь, через n кроків стан системи s^n буде виражатися через початковий стан s^0 формулою $s^n = P^n s^0$. Отже, справедлива формула $P(n) = P^n$.

Теорема Маркова (про граничні ймовірності): якщо існує таке натуральне число n_0 , що всі елементи матриці $P(n_0) = P^{n_0}$ строго додатні, для кожного $j = \overline{1, m}$

існує гранична ймовірність $\lim_{n \rightarrow \infty} P(n) = P^*$

Граничні ймовірності (якщо вони існують) можна знайти із рівнянь:

$$Ps = s, \rightarrow p_{ij}s_j = \delta_{ij}s_j, \quad (p_{ij} - \delta_{ij})s_j = 0.$$

Отже, разом із умовою нормування на компоненти вектора стану задача про відшукування граничного стану зводиться до розв'язання системи рівнянь:

$$(p_{11} - 1)s_1 + p_{12}s_2 + \dots + p_{1n}s_n = 0,$$

$$p_{21}s_1 + (p_{22} - 1)s_2 + \dots + p_{2n}s_n = 0,$$

$$\dots \dots$$

$$p_{n1}s_1 + p_{n2}s_2 + \dots + (p_{nn} - 1)s_n = 0,$$

$$s_1 + s_2 + \dots + s_n = 1.$$

Оскільки дана система перевизначена (вона містить зайве рівняння), із неї можна викреслити одне рівняння. Небажано викреслювати останнє рівняння нормування вектора стану.

Приклад. Нехай (A_1, A_2, A_3) -- точки числової осі із цілочисловими координатами ($x = 1, x = 2, x = 3$). Уявимо собі частинку, що рухається за цими точками таким чином: якщо у деякий момент часу $t = n$ ($n = 0, 1, 2, 3, \dots$) частинка міститься у внутрішній точці A_2 , в наступний момент $t = n + 1$ вона переходить в A_1 із ймовірністю q або в A_3 із ймовірністю $p = 1 - q$; якщо частинка опинилася у лівій межовій точці A_1 , в наступний момент часу із ймовірністю q вона там залишається або із ймовірністю p повертається в A_2 ; якщо ж частинка опинилася у правій межовій точці A_3 , в наступний момент часу вона там залишається із ймовірністю p або повертається в A_2 із ймовірністю q .

- Знайдіть матрицю переходів і побудуйте її граф станів.
- Знайдіть матрицю переходів за два кроки.
- Перевірте існування граничних станів.
- Якщо граничні стани існують, знайдіть їх.

Розв'язання.

- Побудуємо граф станів:

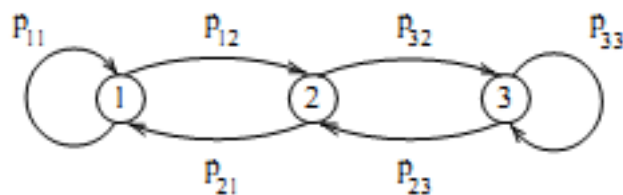


Рис.3.1 Граф станів

Нагадаємо, що початковому стану переходу відповідає другий індекс k , а кінцевому -- перший індекс i матриці p_{ik} . Знаходимо ймовірності переходів p_{ik} за один крок:

$$p_{11} = q, p_{21} = p, p_{22} = 0, p_{23} = q, p_{33} = p.$$

В результаті матриця переходів має вигляд:

$$P = \begin{bmatrix} q & p & 0 \\ p & 0 & q \\ 0 & q & p \end{bmatrix}.$$

В якості перевірки зауважимо, що сума елементів за стовпцями матриці дорівнює одиниці (згідно умові нормування для вектору станів).

б) Для відшукування $P(2)$ обчислюємо P^2 :

$$P(2) = P^2 = \begin{bmatrix} q & p & 0 \\ p & 0 & q \\ 0 & q & p \end{bmatrix} \cdot \begin{bmatrix} q & p & 0 \\ p & 0 & q \\ 0 & q & p \end{bmatrix} = \begin{bmatrix} q^2 + p^2 & qp & qp \\ pq & p^2 + q^2 & pq \\ pq & pq & p^2 + q^2 \end{bmatrix}.$$

в) Оскільки всі елементи P^2 є строго додатні, умова теореми Маркова про граничні ймовірності виконується. Отже, граничні ймовірності (s_1^*, s_2^*, s_3^*) існують.

г) Для відшукування граничних станів розв'яжемо систему:

$$\begin{bmatrix} q-1 & q & 0 \\ p & 0-1 & q \\ 0 & p & p-1 \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

із додатковою умовою $s_1 + s_2 + s_3 = 1$.

$$s_1 = \frac{q^2}{q^2 + pq + p^2}; \quad s_2 = \frac{pq}{q^2 + pq + p^2}; \quad s_3 = \frac{p^2}{q^2 + pq + p^2}.$$

3.2.2 Робота телефонного комутатора

Приклад. Система, що досліджується, (комутатор) може приймати два стани: вона може бути вільною у момент часу із ймовірністю $P_0(t)$ і зайнятою із ймовірністю $P_1(t)$.

Якщо лінія вільна, за проміжок часу Δt на комутатор надходить сигнал із ймовірністю $\alpha \Delta t$ і переводить систему у стан зайнято. Якщо комутатор зайнятий, із ймовірністю $\beta \Delta t$ він обробляє сигнал і переходить у стан вільно. Знайти граничний стан системи при $\alpha = 0,5$, $\beta = 0,3$.

Розв'язання. Припустимо, у момент часу $t + \Delta t$ система була вільною. Це означає, що у попередній момент часу t :

1. Система перебувала у стані вільно $P_0(t)$ і до неї не надійшов сигнал (ймовірність ненадходження сигналу дорівнює $1 - \alpha \Delta t$);

2. Система перебувала у стані зайнято $P_1(t)$, але за проміжок часу Δt запит було оброблено із ймовірністю $\beta \Delta t$.

Тоді за формулою повної ймовірності отримаємо:

$$P_0(t + \Delta t) = P_0(t)(1 - \alpha \Delta t) + P_1(t)\beta \Delta t$$

або

$$P_0(t + \Delta t) - P_0(t) = -P_0(t)\alpha \Delta t + P_1(t)\beta \Delta t.$$

Розділимо обидві частини рівняння на Δt і візьмемо границю $\Delta t \rightarrow 0$:

$$\lim_{\Delta t \rightarrow 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -P_0(t)\alpha + P_1(t)\beta = P'_0(t).$$

Маємо диференціальне рівняння першого порядку

$$dP_0/dt = -P_0\alpha + P_1\beta.$$

Тепер припустимо, що у момент часу $t + \Delta t$ система була зайнята. Це означає, що у попередній момент часу t :

1. Система перебувала у стані вільно (із ймовірністю $P_0(t)$) і до неї надійшов сигнал (ймовірність надходження сигналу дорівнює $\alpha\Delta t$);
2. Система перебувала у стані зайнято (із ймовірністю $P_1(t)$) і за проміжок часу Δt запит так і не було оброблено (ймовірність того, що сигнал не було оброблено, дорівнює $1 - \beta\Delta t$).

Тоді за формулою повної ймовірності отримаємо:

$$P_1(t + \Delta t) - P_1(t) = P_0(t)\alpha\Delta t - P_1(t)\beta\Delta t.$$

Розділимо обидві частини рівняння на Δt й візьмемо границю.

$$dP_1/dt = \alpha P_0 - \beta P_1.$$

За припущення, що у початковий момент часу $t = 0$ система була вільна (із ймовірністю $P_0(0) = 1$ (тоді $P_1(0) = 0$)), отримуємо систему диференціальних рівнянь

$$P_0' = -\alpha P_0 + \beta P_1, \quad P_0(0) = 1,$$

$$P_1' = \alpha P_0 - \beta P_1, \quad P_1(0) = 0.$$

Примітка. Я використовую у якості позначення оператора Лапласа позначення p , що збігається із загальноприйнятим позначенням ймовірності подій p . Прошу звернути на це увагу і вибачити за можливі незручності!

Перетворення за Лапласом даної системи дає

$$(p + \alpha)P_0 - \beta P_1 = 1;$$

$$\alpha P_0 - (p + \beta)P_1 = 0.$$

Зворотне перетворення дає

$$P_0(t) = \frac{\alpha}{\alpha + \beta} (e^{-(\alpha + \beta)t} + 1);$$

$$P_1(t) = \frac{\alpha}{\alpha + \beta} (1 - e^{-(\alpha + \beta)t}).$$

Для конкретних значень $\alpha = 0,5$, $\beta = 0,3$ графік розв'язків набуває вигляду

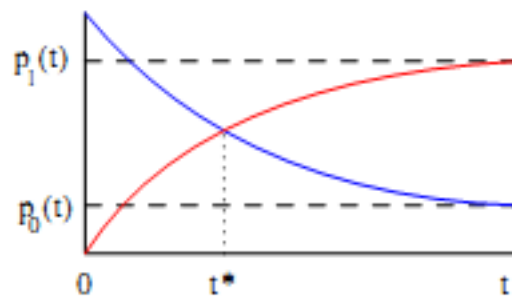


Рис.3.3 Графік розв'язків для комутатора

При $t \rightarrow \infty$ граничні стани системи мають вигляд

$$P_0(t) = \frac{\beta}{\alpha + \beta}; \quad P_1(t) = \frac{\alpha}{\alpha + \beta}.$$

Із графіка видно, що крива ймовірності вільного стану перетинається із кривою завантаженого стану.

Це означає, що після настання деякого часу t^* система перестане справлятися із

потокком заявок. Тому необхідно або збільшити продуктивність системи (зменшити час обслуговування заявки), або зменшити сам потік заявок.

Приклад. Досліджується мікропроцесор, що розрахований на одночасне обслуговування трьох контролерів. Знайти граничний стан системи, якщо ймовірність надходження переривання на мікропроцесор за проміжок часу Δt дорівнює $\alpha\Delta t$. Ймовірність обробки переривання за проміжок часу Δt дорівнює $\beta\Delta t$. ($\alpha = 0,3$, $\beta = 0,5$).

Розв'язання. Позначимо через $P_0(t)$ ймовірність того, що мікропроцесор є вільний (на нього не надходило жодного переривання). Тоді $P_1(t)$ -- ймовірність обробки мікропроцесором одного переривання, $P_2(t)$ -- двох переривань, $P_3(t)$ -- трьох переривань. Запровадимо позначення для таких станів мікропроцесора:

x_0 -- вільний; x_1 -- зайнятий рівно один вхід; x_2 -- зайнято два входи; x_3 -- зайняті всі три входи.

1). Припустимо, що у момент часу $t + \Delta t$ система була вільною. Це означає, що у попередній момент часу t :

1. Система перебувала у стані вільно x_0 (із ймовірністю $P_0(t)$) і до неї не надійшов сигнал (ймовірність ненадходження сигналу дорівнює $1 - \alpha\Delta t$);

2. Система перебувала у стані зайнято x_1 , (із ймовірністю $P_1(t)$), але за проміжок часу Δt запит було оброблено із ймовірністю $\beta\Delta t$.

Аналогічно попередньому прикладу

$$dP_0/dt = -\alpha P_0 + \beta P_1.$$

2) Тепер припустимо, що у момент часу $t + \Delta t$ мікропроцесор був зайнятий обробкою одного переривання. Це означає, що у попередній момент часу t :

1. Система перебувала у стані вільно x_0 (із ймовірністю $P_0(t)$) і до неї надійшов сигнал (ймовірність надходження сигналу дорівнює $\alpha\Delta t$);

2. Система перебувала у стані зайнято x_1 і за проміжок часу Δt система не обробила переривання і до неї не надійшов жодний сигнал (ймовірність $1 - \alpha\Delta t - \beta\Delta t$).

3. Система перебувала у стані x_2 і за проміжок часу Δt оброблено один запит із двох (ймовірність $2\beta\Delta t$).

Тоді за формулою повної ймовірності отримаємо:

$$P_1(t + \Delta t) = P_0(t)\alpha\Delta t + P_1(t)(1 - \alpha\Delta t - \beta\Delta t) + 2\beta\Delta t P_2(t),$$

або

$$P_1(t + \Delta t) - P_1(t) = \alpha\Delta t P_0(t) - (\alpha + \beta)\Delta t P_1(t) + 2\beta\Delta t P_2(t).$$

Після граничного переходу отримуємо

$$dP_1/dt = \alpha P_0 - (\alpha + \beta)P_1 + 2\beta P_2.$$

3) Тепер припустимо, що у момент часу $t + \Delta t$ мікропроцесор перебував у стані x_2 (обробка двох переривань). Це означає, що у попередній момент часу t :

1. Система перебувала у стані x_1 (із ймовірністю $P_1(t)$) і до неї надійшов сигнал (із ймовірністю $\alpha\Delta t$);

2. Система перебувала у стані x_2 (із ймовірністю $P_2(t)$) і за проміжок часу Δt запит так і не було оброблено і не надійшов жодний запит (ймовірність $1 - \alpha\Delta t - 2\beta\Delta t$).

3. Мікропроцесор був завантажений обробкою трьох переривань (із ймовірністю $P_3(t)$) і за проміжок часу Δt обробив одне із трьох переривань. Тобто перейшов із стану

x_3 у стан x_2 із ймовірністю $3\beta\Delta t$.

Тоді за формулою повної ймовірності отримаємо:

$$P_2(t + \Delta t) - P_2(t) = \alpha\Delta t P_1(t) - (\alpha + 2\beta)\Delta t P_2(t) + 3\beta\Delta t P_3(t).$$

$$dP_2/dt = \alpha P_1 - (\alpha + 2\beta)P_2 + 3\beta P_3.$$

4) Тепер припустимо, що у момент часу $(t + \Delta t)$ мікропроцесор перебував у стані x_3 (був сповна завантажений). Це означає, що у попередній момент часу t :

1. Система перебувала у стані x_2 (із ймовірністю $P_2(t)$) і до неї надійшов сигнал (із ймовірністю $\alpha\Delta t$);

2. Система перебувала у стані x_3 (із ймовірністю $P_3(t)$) і за проміжок часу Δt так і не був оброблений жоден із трьох запитів (ймовірність необроблення сигналів дорівнює $1 - 3\beta\Delta t$).

$$P_2(t + \Delta t) = \alpha P_2(t)\Delta t + (1 - 3\beta\Delta t)P_3(t).$$

$$dP_3/dt = \alpha P_2 - 3\beta P_3.$$

За припущення, що у початковий момент часу $t = 0$ система була вільною (із ймовірністю $P_0(0) = 1$) (тоді $P_1(0) = 0$, $P_2(0) = 0$, $P_3(0) = 0$), отримуємо систему диференціальних рівнянь

$$P_0' = -\alpha P_0 + \beta P_1, \quad P_0(0) = 1,$$

$$P_1' = \alpha P_0 - (\alpha + \beta)P_1 + 2\beta P_2, \quad P_1(0) = 0,$$

$$P_2' = \alpha P_1 - (\alpha + 2\beta)P_2 + 3\beta P_3, \quad P_2(0) = 0,$$

$$P_3' = \alpha P_2 - 3\beta P_3, \quad P_3(0) = 0.$$

Перетворення за Лапласом даної системи дає (із урахуванням початкових умов)

$$((p + \alpha)P_0 - \beta P_1 = 1,$$

$$\alpha P_1 - (p + \alpha + \beta)P_1 + 2\beta P_2 = 0,$$

$$\alpha P_1 - (p + \alpha + 2\beta)P_2 + 3\beta P_3 = 0,$$

$$\alpha P_2 - (p + 3\beta)P_3 = 0.$$

Запишемо цю систему у матричному вигляді $AP = b$:

$$A = \begin{bmatrix} \alpha + p & -\beta & 0 & 0 \\ -\alpha & \alpha + \beta + p & -2\beta & 0 \\ 0 & -\alpha & p + \alpha + 2\beta & -3\beta \\ 0 & 0 & -\alpha & p + 3\beta \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Розв'язання цього матричного функціонального рівняння, що залежить від параметрів α, β , отримано у такому вигляді. Характеристичне рівняння має вигляд

$$\Delta = p[p^3 + 3(\alpha + 2\beta)p^2 + (3\alpha^2 + 9\alpha\beta + 11\beta^2)p + (\alpha^3 + 3\alpha^2\beta + 6\alpha\beta^2 + 6\beta^3)].$$

$$P_0(p, \alpha, \beta) = [p^3 + 2(\alpha + 3\beta)p^2 + (\alpha^2 + 4\alpha\beta + 11\beta^2)p + 6\beta^3]/\Delta;$$

$$P_1(p, \alpha, \beta) = \{\alpha[p^2 + (\alpha + 5\beta)p + 6\beta^2]\}/\Delta;$$

$$P_2(p, \alpha, \beta) = \alpha^2(p + 3\beta)/\Delta; \quad P_3(p, \alpha, \beta) = \alpha^3/\Delta.$$

Корені характеристичного рівняння доцільно відшукувати за фіксованих значень параметрів α, β . Позначимо їх: $r_1 = 0$, r_2 , r_3 , r_4 . Тоді після відшукування власних значень (поліосів) характеристичного рівняння розв'язання матричного рівняння можна представити у вигляді сум елементарних складових:

$$P_m(t, \alpha, \beta) = \sum_{k=1}^4 \frac{p_{m,k}}{p + r_k}; \quad m = \overline{0,3}.$$

У просторі оригіналів маємо

$$P_m(t, \alpha, \beta) = \sum_{k=1}^4 p_{mk} e^{-r_k t}; \quad m = \overline{0,3}.$$

для конкретних значень графік розв'язків набуває вигляд 3.4:

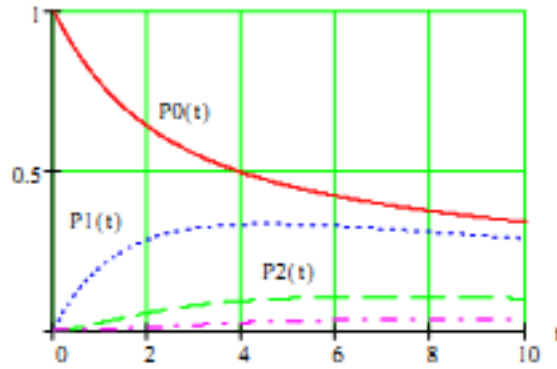


Рис.3.4 Графік розв'язків

Із графіка видно, що криві ймовірності вільного стану перетинаються із кривими завантаженого стану. Це означає, що система справляється із потоком заявок та її продуктивності досить для обслуговування всіх переривань.

Задача. Розглядається робота офісної міні-АТС, що розрахована на обслуговування N абонентів (N -канальна система). Знайти граничний стан системи.

3.2.3 Метод вкладених ланцюгів Маркова

Метод вкладених ланцюгів Маркова доцільно використовувати коли або вхідний потік або потік обслуговування є марківський, а інший є довільний. У цьому випадку випадковий процес у СМО розглядають у завчасно визначених точках регенерації (точках відновлення), що утворюють ланцюг Маркова. Для пуассонівського вхідного потоку точками регенерації є моменти часу, коли чергова процедура обслуговування завершена, а відповідна вимога саме залишає систему. Для часу обслуговування, що розподілений за експоненційним законом, у якості точок регенерації обираються моменти часу, що збігаються із моментами надходження нових запитів. Після цього конкретизуються події, що відбуваються навколо цих точок, і визначається закон розподілу ймовірності кількості подій вхідного пуассонівського потоку або потоку обслуговування за час між сусідніми точками регенерації.

Будується матриця ймовірностей

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots & p_{0n} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1n} \\ p_{20} & p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & & & & \\ p_{n0} & p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \quad (1)$$

де p_{jk} -- ймовірність переходу зі стану j у стан k за час між сусідніми точками регенерації.

За цією матрицею можна скласти систему алгебраїчних рівнянь для ймовірностей станів:

$$p_j = \sum_{k=0}^n p_k \cdot p_{jk}, \quad j = \overline{0, n}, \quad (2)$$

яка доповнюється умовою нормування $\sum_{j=0}^n p_j = 1$.

Для одноканальної СМО з пуассонівським вхідним потоком, який характеризується інтенсивністю λ і функцією розподілу часу обслуговування $B(t)$ $p_{jk} = r_{j=k+1}$, $j > 0$; $p_{0k} = r_k$,

$$r_k = \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^m}{m!} dB(t), \quad m \geq 0, \quad m < 0 \rightarrow r_m = 0$$

матриця перехідних ймовірностей перетворюється на:

$$P = \begin{bmatrix} r_0 & r_1 & r_2 & r_3 & \dots \\ r_0 & r_1 & r_2 & r_3 & \dots \\ 0 & r_0 & r_1 & r_2 & \dots \\ 0 & 0 & r_0 & r_1 & \dots \\ \dots, \end{bmatrix}$$

Розглянемо СМО із груповим вибором запитів за умови, що вхідний потік описується функцією розподілу $A(t)$, а процес вибору запитів описується експоненціальним розподілом із інтенсивністю μ , []. За стан вкладеного ланцюга Маркова оберемо кількість вимог у накопичувачі безпосередньо перед надходженням чергової вимоги. Позначимо p_{ij} ймовірність переходу за час t між двома точками регенерації. За час τ може надійти запит на перепис вмісту накопичувача із ймовірністю $1 - e^{-\mu\tau}$. Тоді процес перейде зі стану s_i у стан s_0 , $i = \overline{0, k}$ із ймовірністю

$$p_{i0} = \int_0^{\infty} (1 - e^{-\mu t}) dA(t) = 1 - \alpha(\mu),$$

де $\alpha(\mu)$ -- перетворення Лапласа--Стілтєса від $A(t)$ за параметром μ .

За час τ пристрій, що обслуговує, може не вибрати вмісту накопичувача, ймовірність цього -- $e^{-\mu\tau}$. Процес перейде зі стану i у стан $i+1$, $i = \overline{0, k-1}$ із ймовірністю

$$p_{i, i+1} = \int_0^{\infty} (e^{-\mu t}) dA(t) = \alpha(\mu).$$

Якщо процес перебуває у стані k , він залишиться у цьому стані із ймовірністю $p_{kk} = \alpha(\mu)$, тобто запит, що недавно надійшов, буде втрачено. Матриця перехідних ймовірностей має вигляд:

$$P = \begin{bmatrix} 1 - \alpha(\mu) & \alpha(\mu) & 0 & \dots & 0 \\ 1 - \alpha(\mu) & 0 & \alpha(\mu) & \dots & 0 \\ \dots & & & & \\ 1 - \alpha(\mu) & 0 & 0 & \dots & \alpha(\mu) \end{bmatrix}.$$

Для ланцюга, що досліджується, існує стаціонарний стан із ймовірностями p_i ,

$i = \overline{0, k}$ того, що у накопичувачі перед моментом регенерації було $i = \overline{0, k}$ запитів. Для цих ймовірностей справедлива система рівнянь

$$p_j = \sum_{i=0}^k p_i \cdot p_{ij}, \quad j = \overline{0, k}, \quad (3)$$

з умовою нормування

$$\sum_{j=0}^n p_j = 1. \quad (4)$$

Розв'язок системи рівнянь (4) має вигляд

$$p_j = [1 - \alpha(\mu)][\alpha(\mu)]^j, \quad 0 \leq j < k; \quad p_k = [\alpha(\mu)]^k.$$

Середня кількість запитів у накопичувачі:

$$L_q = \sum_{j=1}^k j p_j = \alpha(\mu) \{1 - [\alpha(\mu)]^k\} / [1 - \alpha(\mu)].$$

3.2.4 СМО із очікуванням

Розглянемо ситуацію, в якій заявка, що застає всі канали зайнятими, стає у чергу й очікує, поки не звільниться будь-який канал. Вочевидь, для постановки заявки у чергу мають бути виділені відповідні ресурси (обсяг кеш-пам'яті процесора, кількість місць у приймальні начальника тощо). Якщо ресурси (місця) для очікування заповнено, заявка у чергу не стає.

Позначимо $\alpha \Delta t$ -- ймовірність надходження заявки у систему; $\beta \Delta t$ -- ймовірність обслуговування заявки; $\gamma \Delta t$ -- ймовірність уходу заявки із черги.

У загальному випадку n -канальна система із k -розрядною пам'яттю (пам'ять, що утримує k заявок) може приймати такі стани:

x_0 -- всі канали вільні; x_1 -- один канал зайнято; x_2 -- два канали зайняті;

x_n -- всі n каналів зайняті;

x_{n+1} -- всі канали зайняті й одна заявка у черзі; x_{n+2} -- зайнято два місця у черзі;

x_{n+k} -- зайняті всі n каналів і k місць у черзі.

Система диференціальних рівнянь у даному випадку приймає вигляд

$$\begin{aligned} P'_0 &= -\alpha P_0 + \beta P_1, \\ P'_1 &= \alpha P_0 - (\alpha + \beta) P_1 + 2\beta P_2, \\ P'_2 &= \alpha P_1 - (\alpha + 2\beta) P_2 + 3\beta P_3, \\ &\dots \\ P'_{n-1} &= \alpha P_{n-2} - (\alpha + (n-1)\beta) P_{n-1} + n\beta P_n, \\ P'_n &= \alpha P_{n-1} - (\alpha + n\beta) P_n + (n\beta + \gamma) P_{n+1}, \\ P'_{n+1} &= \alpha P_n - (\alpha + n\beta + \gamma) P_{n+1} + (n\beta + 2\gamma) P_{n+2}, \\ P'_{n+2} &= \alpha P_{n+1} - (\alpha + n\beta + 2\gamma) P_{n+2} + (n\beta + 3\gamma) P_{n+3}, \\ &\dots \\ P'_{n+k} &= \alpha P_{n+k-1} - (n\beta + k\gamma) P_{n+k}. \end{aligned}$$

Початкові умови: $P_0(0) = 1, P_1(0) = P_2(0) = \dots = P_{n+k}(0) = 0$.

Приклад. На трьохканальний комутатор із 2-х розрядною пам'яттю надходить потік заявок. Ймовірність надходження заявки на комутатор за проміжок часу Δt дорівнює $\alpha\Delta t$; ймовірність обслуговування $\beta\Delta t$; ймовірність уходу заявки із черги $\gamma\Delta t$.

Знайти граничний стан системи за $\alpha = 0,5$, $\beta = 0,3$.

Розв'язання. Система може приймати одне із таких можливих значень:

x_0 -- жоден канал не зайнятий; x_1 -- зайнятий рівно один канал; x_2 -- зайнято два канали; x_3 -- зайняті три канали; x_4 -- зайняті всі канали, й одна заявка чекає черги; x_5 -- зайняті всі канали, і дві заявки стоять у черзі.

Тоді система диференціальних рівнянь для відповідних ймовірностей має вигляд

$$\begin{aligned}P'_0 &= -\alpha P_0 + \beta P_1, \\P'_1 &= \alpha P_0 - (\alpha + \beta) P_1 + 2\beta P_2, \\P'_2 &= \alpha P_1 - (\alpha + 2\beta) P_2 + 3\beta P_3, \\P'_3 &= \alpha P_2 - (\alpha + 3\beta) P_3 + (3\beta + \gamma) P_4, \\P'_4 &= \alpha P_3 - (\alpha + 3\beta + \gamma) P_4 + (3\beta + \gamma) P_5, \\P'_5 &= \alpha P_4 - (\alpha + 3\beta + 2\gamma) P_5.\end{aligned}$$

Початкові умови: $P_0(0) = 1$, $P_1(0) = P_2(0) = P_3(0) = P_4(0) = P_5(0) = 0$.

Перетворення за Лапласом і запис у матричному вигляді дає $AP = b$, де $b = [1\ 0\ 0\ 0\ 0\ 0]^T$.

$$\begin{bmatrix} p + \alpha & -\beta & 0 & 0 & 0 & 0 \\ -\alpha & \alpha + \beta + p & -2\beta & 0 & 0 & 0 \\ 0 & -\alpha & \alpha + 2\beta + p & -3\beta & 0 & 0 \\ 0 & 0 & \alpha & \alpha + 3\beta + p & -(3\beta + \gamma) & 0 \\ 0 & 0 & 0 & \alpha & \alpha + 3\beta + \gamma + p & -(3\beta + 2\gamma) \\ 0 & 0 & 0 & 0 & \alpha & \alpha + 3\beta + 2\gamma + p \end{bmatrix}.$$

Знайдемо її розв'язання.

Характеристичний поліном отримано у вигляді

$$\Delta = p^6 + (6\alpha + 12\beta + 3\gamma)p^5 + (15\alpha + 56\beta^2 + 2\gamma^2 + 6\alpha(11\beta + 3\gamma) + 27\beta\gamma)p^4$$

$$+ [4\alpha^2(5\alpha + 36\beta + 10,5\gamma) + 3\beta^2(83\alpha + 4\beta + 29\gamma) + 12\beta\gamma(11\alpha + \gamma)]p^3$$

$$+ [3\alpha^3(5\alpha + 52\beta + 16\gamma) + 9\beta^3(45\alpha + 15\beta + 13\gamma) + \alpha\beta\gamma(42\gamma + 288\beta + 231\alpha)$$

$$+ 6\alpha^2(3\gamma^2 + 83\beta^2) + 22\beta^2\gamma^2]p^2 + [3\alpha^4(3\alpha + 26\beta + 9\gamma) + 18\beta^4(11\alpha + 3\beta + 3\gamma) +$$

$$\alpha^2\beta^2(327\beta + 287\alpha + 261\gamma) + \alpha\beta\gamma(174\alpha + 174\beta + 32\beta\gamma + 42\alpha\gamma) + 2\gamma^2(6\alpha^3 + 6\beta^3)]p$$

$$+ [54\alpha^3\beta^2(\beta + \gamma) + 6\alpha^3\beta(3\alpha^2 + 8\alpha\gamma + 2\beta\gamma^2) + \alpha^3(\alpha^3 + 6\alpha^2\gamma + 4\alpha\gamma^2)].$$

Вирази для шуканих функцій у просторі зображень отримано у такому вигляді.

$$P_0(\alpha, \beta, \gamma) = \{ p^5 + (5\alpha + 12\beta + 3\gamma)p^4 +$$

$$+ [10\alpha^2 + 5\alpha\beta + 12\gamma^2 + 56\beta^2 + 27\beta\gamma + 15\alpha\gamma]p^3 +$$

$$\begin{aligned}
& +[\alpha^2(10\alpha + 93\beta + 27\gamma) + 3\beta^2(42\beta + 69\alpha + 29\gamma) + 4\gamma^2(2\alpha + 3\beta) + 108\alpha\beta\gamma)]p^2 + \\
& +\alpha^3(5\alpha + 69\beta + \gamma) + 3\beta^3(45\beta + 103\alpha + 39\gamma) + \alpha^2(243\beta^2 + 10\gamma^2 + 135\beta\gamma) + 22\beta^2\gamma^2 + \\
& +\alpha\beta\gamma(32\gamma + 135\alpha + 225\beta)]p + [\alpha^4(\alpha + 19\beta + 6\gamma) + 18\beta^4(3\beta + 3\gamma + 4\alpha) + 2\alpha^3(2\gamma^2 + 27\beta\gamma \\
& + 16\beta^2) + 4\beta^3(39\alpha + 3\gamma^2) + \alpha\beta\gamma(16\alpha\gamma + 120\beta^2 + 114\alpha\beta + 20\beta\gamma)]\}/\Delta;
\end{aligned}$$

$$\begin{aligned}
P_1(\alpha, \beta, \gamma) = & \alpha\{[p^4 + (4\alpha + 11\beta + 3\gamma)p^3 + (6\alpha^2 + 42\alpha\beta + 45\beta^2 + 24\beta\gamma + 12\alpha\gamma)p^2 + \\
& +[\alpha^2(4\alpha + 51\beta + 15\gamma) + \beta(81\beta + 138\alpha^3\gamma + \gamma^2(6\alpha + 10\beta^2) + 78\alpha\beta\gamma)]p + \alpha^3(\alpha + 20\beta + 6\gamma) + \\
& +\beta^3(54\beta + 144\alpha + 54\gamma) + \alpha^2(102\beta^2 + 60\beta\gamma + 4\gamma^2) + 20\alpha\beta\gamma(6\beta + \gamma) + 18\beta^2\gamma^2]\}/\Delta;
\end{aligned}$$

$$\begin{aligned}
P_2(\alpha, \beta, \gamma) = & \alpha^2\{p^3 + (3\alpha + 9\beta + 3\gamma)p^2 + (3\alpha^2 + 27\beta^2 + 2\gamma^2 + 24\alpha\beta + 9\alpha\gamma + 18\beta\gamma)p + \\
& +[\alpha^2(4\alpha + 15\beta + 6\gamma) + \beta^2(27\beta + 45\alpha + 27\gamma) + 33\alpha\beta\gamma]\}/\Delta;
\end{aligned}$$

$$P_3(\alpha, \beta, \gamma) = -\alpha^3[p^2 + (2\alpha + 6\beta + 3\gamma)p + (\alpha^2 + 9\beta^2 + 2\gamma^2 + 9\alpha\beta + 5\alpha\gamma + 9\beta\gamma)]/\Delta;$$

$$P_4(\alpha, \beta, \gamma) = \alpha^4[p + (\alpha_3\beta + 2\gamma)]/\Delta; \quad P_5(\alpha, \beta, \gamma) = -\alpha^5/\Delta.$$

За аналогією із попереднім прикладом позначимо власні значення системи (корені характеристичного рівняння ($\Delta = 0$) через r_k , ($k = \overline{1, 6}$).

У просторі оригіналів маємо

$$P_m(t, \alpha, \beta) = \sum_{k=1}^6 p_{mk} e^{-r_k t}, \quad m = \overline{0, 5}.$$

Задача. Розглядається робота офісної міні-АТС, що розрахована на одночасне обслуговування n абонентів (n -канальна система), яка може тримати у пам'яті (на черзі) k викликів. Знайти граничний стан системи.

N	n	α	β	N	n	α	β	N	n	α	β
1	5	0.21	0.36	11	5	0.11	0.31	21	7	0.21	0.19
2	6	0.22	0.37	12	6	0.12	0.32	22	6	0.22	0.20
3	7	0.23	0.38	13	5	0.13	0.33	23	7	0.23	0.21
4	6	0.24	0.39	14	6	0.14	0.34	24	5	0.24	0.22
5	5	0.25	0.40	15	7	0.15	0.35	25	6	0.25	0.23
6	6	0.26	0.41	16	6	0.16	0.49	26	7	0.26	0.24
7	5	0.27	0.42	17	5	0.17	0.50	27	4	0.27	0.25
8	6	0.28	0.43	18	6	0.18	0.21	28	6	0.28	0.46
9	7	0.29	0.44	19	7	0.19	0.33	29	5	0.29	0.47
10	6	0.30	0.45	20	6	0.20	0.34	30	7	0.30	0.48

Рис.3.5 Варіанти завдань

3.3 Контрольні запитання

1. З яких елементів складається система масового обслуговування?
2. Якими характеристиками оцінюється система масового обслуговування?
3. Як описується потік подій системи масового обслуговування?
4. Як описується процес обслуговування СМО?
5. Якими методами досліджуються СМО?
6. Які закони розподілу ймовірностей використовуються у СМО?
7. У чому полягає сутність диференційного методу дослідження СМО?

Розділ 4 Планування експериментів

4.1 Вступ

Математичне планування експерименту -- відносно новий підхід до організації і проведення експериментальних досліджень багатофакторних об'єктів. За статистичного підходу до вивчення багатофакторних об'єктів природа цих об'єктів на має суттєвого значення. Об'єкт подається як "чорна скринька", що характеризується набором вхідних параметрів $x_1, x_2, \dots, x_k$ та вихідних параметрів $y_1, \dots, y_m$, залежність між якими підлягає визначенню за даними експерименту. Поняття "чорної скриньки", запроваджене у кібернетиці, відповідає функціональній системі, внутрішня структура якої невідома. Вхідні параметри відповідають можливим способам впливу на об'єкт і називаються *факторами*. Це можуть бути фізичні, хімічні, біологічні тощо впливи, види ліків, їхні дози і спосіб їхнього введення, різні стани (вік, порода, сорт тощо).

Вихідні параметри характеризують результат впливів і називаються *відгуками, параметрами оптимізації*. Це можуть бути врожай біомаси, вихід клітин, вміст гемоглобіну у крові, показники ефективності лікування і інші ознаки, що характеризують результат впливу факторів.

Задача статистичного дослідження багатофакторного об'єкта зводиться до визначення кількісної залежності вихідних параметрів від вхідних факторів, відокремлення основних факторів, що суттєво впливають на відгук, і підбору режимів, які забезпечують оптимальні значення вихідних показників.

Залежність вхідних і вихідних параметрів $y_i = f_i(x)$ не є строго детермінована, оскільки на результати експерименту впливають збурення неконтрольованих факторів. У медицині -- це похибки вимірювань або неконтрольований вплив зовнішнього середовища; у біології -- варіації від окремого організму до іншого або варіація фізіологічних станів одного й того самого організму. Ця залежність, що визначається за даними експерименту, називається *функцією відгуку, або статистичною моделлю об'єкту*.

Для визначення моделі існує два способи накопичення статистичної інформації: "пасивне" спостереження і "активний", спланований спеціальним чином експеримент. Пасивний експеримент ґрунтується на обробці інформації, отриманої у процесі природної поведінки об'єкта, і часто є єдиним можливим, оскільки постановка активного експерименту (наприклад, у медико-біологічних дослідженнях) не завжди можлива.

Потреби практики призвели до розвитку математичного планування експерименту, що ґрунтується на активних експериментах за завчасно складених програмах. Планувати експеримент -- значить вибрати оптимальну схему експерименту, яка надає можливість отримати необхідну інформацію про об'єкт за мінімальну кількість дослідів.

При створенні математичної моделі експерименту і виборі методу планування суттєву роль відіграє *імітація експерименту за допомогою ЕОМ*. Імітація полягає у створенні алгоритму, який забезпечує результати, аналогічні результатам експерименту. Так, коли розглядається задача відновлення невідомої функції за результатами вимірювання її значень, корисно обирати похому функцію, обчислювати її значення на ЕОМ і додавати похибку, що імітує процес вимірювання. Похибку генерують за допомогою датчика псевдовипадкових чисел. Потім можна випробувати на цій моделі різні засоби планування експерименту та оброблення даних.

Отже, планування експерименту передбачає наявність математичної моделі і

критерію оптимальності. Підґрунтям багатьох розділів планування експериментів є статистична теорія лінійного регресійного аналізу.

Планування експерименту надає можливість:

1. Зменшити похибку експерименту і запобігти впливу факторів, що заважають;
2. Скоротити кількість дослідів і заданою точністю отримати відповідь на питання;
3. Отримати для об'єктів, що досліджуються, математичні моделі із певними оптимальними властивостями;

4. Приймати рішення на ґрунті чітких формалізованих правил.

4.2 Характеристика основних задач і методів многофакторного експерименту

4.2.1 Вступ

Мета планування експерименту -- отримання більшої кількості інформації за менших витрат.

Основними етапами дослідницької роботи є планування, експеримент та аналіз. Експеримент містить:

- 1) постановку задачі;
- 2) вибір відгуку (залежної змінної);
- 3) вибір факторів, що варіюються;
- 4) вибір рівнів факторів.

Етап планування визначає:

- 1) необхідну кількість спостережень;
- 2) порядок виконання експерименту;
- 3) метод рандомизації;
- 4) математичну модель для опису експерименту.

Етап аналізу -- це:

- 1) збір та обробка даних;
- 2) обчислення статистик для перевірки гіпотез;
- 3) інтерпретація результатів.

4.2.2 Вибір факторів

При виборі області експерименту перш за все треба оцінити межі областей визначення факторів. При цьому мають враховуватися обмеження кількох типів. 1-й -- принципові обмеження для значень факторів, які не можна порушувати за жодних обставин. 2-й тип -- обмеження, що пов'язані із техніко-економічними міркуваннями (наприклад, коштовність сировини, дефіцитністю окремих компонентів, часом ведення процесу). 3-й тип обмежень визначається конкретними умовами проведення процесу, наприклад, апаратурою, технологією тощо.

Цю інформацію називають *апріорною*.

4.2.3 Вибір основного рівня та інтервалів варіювання

Найкращим умовам, що визначаються за допомогою апріорної інформації, відповідає комбінація рівнів факторів. Кожна комбінація є многовимірною точкою у факторному просторі, яку можна розглядати як початкову точку (основний рівень) для

побудови плану експерименту. Побудова плану експерименту зводиться до вибору експериментальних точок, симетричних відносно нульового рівня.

У загальному випадку вибір основного рівня виконується відповідно до схеми, що наведена на рис.

Далі для кожного фактору треба вибрати дві рівні, на яких він буде варіюватися.

Інтервалом варіювання факторів називають число (своє для кожного фактора), додавання якого до основного рівня дає верхній, а віднімання -- нижній рівень фактору. Для спрощення запису умов експерименту та обробки експериментальних даних масштаби по осям вибираються так, щоб верхній рівень відповідав +1, а нижній -- -1, а основний -- нулю. Для факторів із неперервною областю визначення це завжди можна зробити за допомогою перетворення

$$x_j = \frac{\tilde{x}_j - \tilde{x}_{jk}}{I_j},$$

де x_j -- кодоване значення фактору; $\tilde{x}_j$ -- натуральне значення фактору; $\tilde{x}_{jk}$ -- натуральне значення основного рівня; I_j -- інтервал варіювання; k -- номер фактору.

4.2.4 Задачі класифікації

У рамках планування експерименту можна сформулювати задачі пошуку оптимальних умов культивування мікроорганізмів і клітинних систем, оптимізації процесів біотехнології, виявлення основних факторів, що визначають поведінку медико-біологічних об'єктів і запобігання впливу факторів, що заважають, відшукування адекватних моделей для кількісного опису многофакторних систем і об'єктів, пошук і синтез біологічно активних препаратів у задачах комплексної терапії, пошук раціональних схем обслуговування при променевої терапії, програми масових вакцинацій і оцінку параметрів математичних моделей епідемічних процесів тощо.

Всю різноманітність задач можна класифікувати у такі основні групи:

1. Задачі порівняння варіантів, що досліджуються із максимальною точністю і виключення впливу факторів, що заважають (джерел неоднорідності) на результати цих порівнянь.
2. Задачі математичного опису многофакторного об'єкту (МФО) -- отримання адекватних моделей, що описують поведінку об'єкта.
3. Задачі екстремального експерименту -- відшукування оптимальних режимів функціонування МФО.
4. Задачі експерименту, що відсіює -- виявлення найсуттєвіших факторів МФО та відсіювання несуттєвих.
5. Задачі оцінювання і уточнення параметрів моделей, що описують механізм явищ, а також задачі дискримінації таких моделей.

Розрізняють такі етапи процесу дослідження медико-біологічного об'єкту:

1. Попереднє ознайомлення із об'єктом, його структурою та особливостями, вибір вихідних показників, факторів, області дослідження і рівнів варіювання факторів тощо.
2. Виконання попередніх дослідів, уточнення задачі дослідження, експериментальне вивчення характеристик параметрів, точності методів вимірювання тощо.

Іноді корисно проводити попередній кластерний аналіз наявної інформації з метою виділення однорідних груп (наприклад, розділити популяцію тварин на групи за віком, статтю тощо), і подальші експерименти виконувати для кожної однорідної групи окремо.

Для вказаних цілей можна використовувати методи многовимірної статистики типу факторного аналізу і методу головних компонент. Сумісне їх застосування із методами планування експерименту надає можливість розробляти ефективні процедури для вирішення задач відокремлення інформативних ознак, класифікації та ідентифікації об'єктів у просторі цих ознак тощо і тим самим отримати максимум інформації із наявних даних.

3. За результатами цих етапів формулюються вимоги до математичної моделі, обирається метод її отримання, конкретизуються задачі дослідження ті етапність їх вирішення. Наприклад, при дослідженні сумісної дії кількох ліків, за яких спостерігається бажаний терапевтичний ефект. Потім підбирають дози ліків для отримання максимального ефекту, а потім переходять до клінічних експериментів. Здійснюється збір експериментальної інформації (у режимі пасивного спостереження або методом активного експерименту) і побудова шуканої моделі.

4. Дослідження і аналіз отриманої математичної моделі: перевірка її адекватності, вивчення властивостей, виділення областей для подальшого дослідження або використання моделі для відшукування оптимального режиму (відшукування оптимальних у відповідності до обраного критерію комбінацій факторів, що вимірюються).

5. Інтерпретація моделі, використання її для висування гіпотез про механізм явища або для оптимізації процесу.

4.3 План експерименту

Нехай X -- деяка компактна підмножина евклідова простору R^n . Будемо називати її *множиною планування*. Точки множини x називають умовами планування вимірювань. Розглянемо схему проведення вимірювань, згідно з якою результат вимірювань $y(x_i)$ представляється у вигляді

$$y_j = y(x_j) = \sum_{i=1}^m \theta_i f_i(x_j) + \varepsilon_j, \quad (1)$$

де $\theta_1, \dots, \theta_m$ -- невідомі параметри; $f_1(x), \dots, f_m(x)$ -- відомі функції на X , які називають *базисними функціями*; $\varepsilon_j = \varepsilon(x_j)$ -- випадкові похибки, що задовольняють умови: а) центрованості ($M\varepsilon_j = 0$), б) некорельованості ($M\varepsilon_i \varepsilon_j = 0$ за різних вимірювань у точках $x_i, x_j \in X$, $i \neq j$), в) рівноточності ($M\varepsilon_j^2 = \sigma^2 < \infty$ для будь-якої точки $x_i \in X$).

У схемі вимірювань (1) функцією регресії є

$$\eta(x) = M\{y(x)\} = \sum_{i=1}^m \theta_i f_i(x) = \mathbf{q}f(x), \quad (2)$$

де $\mathbf{q}$ -- вектор невідомих параметрів, $f(x)$ -- вектор базисних функцій.

Якщо вимірювання за схемою (1) здійснюються у точках $x_1, \dots, x_N \in X$, отримуємо класичну лінійну регресійну модель $(Y, F\mathbf{q}, \sigma^2 E_N)$, де

$$Y = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix}; \quad F = \begin{bmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & & \\ f_1(x_N) & \dots & f_m(x_N) \end{bmatrix} \quad (3)$$

Інша форма запису моделі $(Y, F\mathbf{q}, \sigma^2 E_N)$ є

$$Y = F\mathbf{q} + s, \quad s \sim (0, \sigma^2 E_N). \quad (4)$$

Точність оцінювання параметрів $\mathbf{q}$ залежить від матриці F або від вибору точок x_i , якщо покласти $x_{ij} = f_i(x_j)$. Це надає можливість ставити задачу про оптимальний вибір точок -- планування експерименту. Набір цих точок називають *планом експерименту*. Планом експерименту називають також імовірнісну міру, що зосереджена у цих точках із рівними вагами:

$$\xi = \xi_n = \begin{Bmatrix} x_1 & \dots & x_N \\ 1/N & \dots & 1/N \end{Bmatrix}. \quad (5)$$

Плани (5) називають дискретними планами. Множина всіляких дискретних планів (5) позначається Ξ_n . Якщо поміж точок $x_1, \dots, x_N$ є точки, що повторюються, цей план зручно записувати у вигляді

$$\xi = \xi_n = \begin{Bmatrix} x_1 & \dots & x_n \\ p_1 & \dots & p_n \end{Bmatrix}, \quad (6)$$

де

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad p_i = r_i/N, \quad i = \overline{1, n}, \quad (7)$$

r_i -- число повторень i -ї точки плану (6) у плані (5).

Неперервним планом називають план вигляду (5), для якого обмеження (6) послаблено: $p_i \geq 0, \quad \sum_{i=1}^n p_i = 1$. Якщо усі точки x_j ($j = \overline{1, n}$) у плані різні, говорять, що план ξ зосереджений в n точках.

На відміну від множини дискретних планів Ξ_N множина неперервних планів є опукла, тобто для $\forall n \in [0, 1]$ і будь-яких планів ξ_1, ξ_2 опукла комбінація цих планів $(1-\alpha)\xi_1 + \alpha\xi_2$ також є неперервний план.

Інформаційна і дисперсійна матриці

Інформаційною матрицею (Фішера) для моделі (2) називається матриця $A = F^T F$. Якщо план експерименту має вигляд (6) і задовольняє обмеження (6),

$$A = \sum_{i=1}^n r_i f(x_i) f^T(x_i) = N \sum_{i=1}^n p_i f(x_i) f^T(x_i). \quad (8)$$

Інформаційна матриця (8) відіграє визначальну роль при побудові найкращих лінійних оцінок невідомих параметрів $\mathbf{q}$ моделі

$$\mathbf{h}(x, \mathbf{q}) = \theta_0 + \sum_{i=1}^n \theta_i x_i + \sum_{i,j=1, i < j}^n \theta_{ij} x_i x_j + \sum_{i=1}^n \theta_{i,i} x_i^2 + \dots \quad (9)$$

Зокрема, якщо матриця $F^T F$ не вироджена, НЛН-оценка $\hat{\mathbf{q}}$ параметрів $\mathbf{q}$ має вигляд

$$\hat{\mathbf{q}} = (F^T F)^{-1} F^T Y, \quad (10)$$

а дисперсійна матриця цієї оцінки дорівнює

$$\mathcal{D}\hat{\mathbf{q}} = \sigma^2 (F^T F)^{-1} = \sigma^2 C. \quad (11)$$

4.4 Плани першого порядку

Припустимо, що функція регресії -- поліном першого степеню від m змінних $x_1, \dots, x_m$

$$\eta(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_m x_m, \quad (12)$$

а похибки вимірювань центровані, некорельовані і мають однакову дисперсію σ^2 .

Інформаційні і дисперсійні матриці мають розмір $(m+1) \times (m+1)$.

Дискретний план

$$\xi_N = \begin{Bmatrix} x_1 & \dots & x_N \\ 1/N & \dots & 1/N \end{Bmatrix}, \quad (13)$$

який забезпечує незміщені МНК-оцінки параметрів $\theta_0, \theta_1, \dots, \theta_m$ функції регресії (12), будемо називати *планом першого порядку*.

Плани першого порядку широко застосовуються у практиці експериментальних досліджень.

План (13) зручно представляти у вигляді матриці

$$D = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & & & \\ x_{N1} & x_{N2} & \dots & x_{Nm} \end{bmatrix} \quad (14)$$

де x_{ji} ($i = \overline{1, m}$) -- координати точки x_j ($j = \overline{1, N}$). Матриця плану має вигляд

$$F = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1m} \\ x_{20} & x_{21} & \dots & x_{2m} \\ \dots & & & \\ x_{N0} & x_{N1} & \dots & x_{Nm} \end{bmatrix} \quad (15)$$

де $x_{j0} = 1$.

Оскільки план (13) вважається невиродженим,

$$N \geq m+1, \quad \text{rank} D = m, \quad \text{rank} F = m+1.$$

Згідно визначенню план першого порядку ортогональний, якщо стовпці матриці F ; A ; попарно ортогональні, тобто

$$\sum_{l=1}^N x_{li} x_{lj} = 0, \quad i, j = \overline{0, m}, \quad i \neq j. \quad (16)$$

Для ортогональних планів першого порядку виконується умова симетрії

$$\sum_{l=1}^N x_{li} = 0, \quad i = \overline{1, m}, \quad (17)$$

що випливає із (5) при $j = 0$.

На двох прикладах подивимося як впливає вибір плану на дисперсії $D\hat{\theta}_j$ МНК-оцінок $\hat{\theta}_j$ невідомих параметрів θ_j ($j = \overline{1, m}$).

Приклад 1. Нехай $N = 6$, $m = 2$. Розглянемо два плани $\xi_6^{(1)}$ та $\xi_6^{(2)}$ вигляду (14), яким відповідають матриці

$$F_1 = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & -\sqrt{3} \\ 1 & -1 & 0 \\ 1 & 1 & -\sqrt{3} \\ 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = (x_{ij}^{(1)}), \quad F_2 = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = (x_{ij}^{(2)}).$$

План $\xi_6^{(1)}$ ортогональний план першого порядку і

$$\sum_{i=1}^6 (x_{ij}^{(1)})^2 = 6, \quad j = 0, 1, 2.$$

Тому дисперсії МНК-оцінок параметрів θ_j дорівнюють $\sigma^2/6$ і (у силу ортогональності $\xi_6^{(1)}$) ці оцінки некорельовані.

Для плану $\xi_6^{(2)}$ також

$$\sum_{i=1}^6 (x_{ij}^{(2)})^2 = 6, \quad j = 0, 1, 2.$$

Але цей план не є ортогональним, оскільки дисперсійна матриця дорівнює

$$\sigma^2 (F_2^T F_2)^{-1} = \sigma^2 \begin{bmatrix} 1/4 & 1/8 & 1/8 \\ 1/8 & 1/4 & 1/8 \\ 1/8 & 1/8 & 1/4 \end{bmatrix}.$$

Отже, дисперсії усіх трьох МНК-оцінок дорівнюють $\sigma^2/4$, тобто більше за план $\xi_6^{(1)}$. Крім того, МНК-оцінки при використанні плану $\xi_6^{(2)}$ отримуються корельованими.

Приклад 2. Нехай $m = 3$, $N = 4$.

$$F_1 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}.$$

План $\xi_4^{(1)}$, що відповідає матриці F_1 , є ортогональний план першого порядку, за використання якого отримуємо $D\hat{\theta}_j = \sigma^2/4$ ($j = 0, 1, 2, 3$). Легко перевірити, що за використання неортогонального плану $\xi_4^{(2)}$, що відповідає матриці F_2 ,

$$D\hat{\theta}_0 = \sigma^2, \quad D\hat{\theta}_j = \sigma^2/2, \quad j = 1, 2, 3,$$

й, отже, план $\xi_4^{(2)}$ суттєво гірший за план $\xi_4^{(1)}$ з точки зору величин дисперсій МНК-оцінок невідомих параметрів.

4.5 Планування експерименту для нелінійних моделей

4.5.1 Нелінійний регресійний аналіз

Розглянемо схему нелінійного експерименту. Нехай у точках x_j , $j = \overline{1, N}$, множини

планування X спостерігаються значення випадкових величин $y_j = y(x_j)$, що подаються у вигляді

$$y_j = \eta(x_j, \theta) + \varepsilon_j, \quad j = \overline{1, N}, \quad (18)$$

де $\eta(x, \theta)$ -- функція регресії на $X \times \Omega$, що задана із точністю до невідомих параметрів $\theta = (\theta_1, \dots, \theta_m)^T$ із деякої параметричної множини $\Omega \subset R^m$, ε_j -- випадкові похибки вимірювань. Вважається, що похибки вимірювань ε_j центровані ($M[\varepsilon_j] = 0$) і не корельовані ($D[\varepsilon_j] = [\varepsilon_j^2] = \sigma^2 = 0$), причому значення σ^2 може бути невідомим.

У разі коли функція регресії $\eta(x, \theta)$ лінійно залежить від невідомих параметрів θ , тобто

$$\eta(x, \theta) = \theta^T f(x),$$

де $f(x) = [f_1(x), \dots, f_m(x)]^T$ -- вектор відомих (базових) функцій, схема регресійного експерименту є схема лінійної регресії.

Нелінійні регресійні моделі у практиці трапляються досить часто.

а) нелінійність виникає із сутності явища, для опису якого слугує модель;

б) додаткова інформація про істинний характер залежності часто надає можливість вибрати досить точну (тобто таку, що має малу систематичну похибку) нелінійну модель із числом параметрів, суттєво меншим, ніж в аналогічній лінійній моделі;

в) лінеаризація моделі часто несе набагато більше втрат, ніж вигод.

Прикладами нелінійних функцій регресії можуть слугувати такі:

1. Функція регресії

$$\eta(x, \theta) = \sum_{i=1}^m \theta_i e^{-\theta_{i+m} x}$$

виникає у результаті розв'язання систем лінійних диференційних рівнянь і широко використовується при вирішенні, зокрема, кінетичних задач.

2. Функція регресії

$$\eta(x, \theta) = \sum_{i=1}^m \theta_i / [(x - \theta_{i+m})^2 + \theta_{i+2m}]$$

описує резонансні явища і використовується при обробці результатів спектрального аналізу.

3. Функція регресії

$$\eta(x, \theta) = [(x_1 - \theta_1) / ((x_2 - \theta_2))]$$

використовується у задачах слідкування за рухомими об'єктами.

4. Якщо відомо, що функція регресії унімодальна (тобто має один локальний максимум), досить гладка, й існує таке θ_1 , що $\eta(x) = 0$ за $x \leq \theta_1$ і $\lim_{x \rightarrow \infty} \eta(x) = 0$, але невідомий механізм, що призводить до такої залежності, у якості функції регресії можна обирати

$$\eta(x, \theta) = \theta_2 (x - \theta_1)^{\theta_3} e^{-\theta_4 (x - \theta_1)^{\theta_5}}, \quad x \geq 0, \quad \theta_i \geq 0, \quad i = 2, 3, 4, 5.$$

Метод найменших квадратів -- один із найрозповсюдженіших методів оцінювання параметрів нелінійних регресійних моделей.

Оцінка МНК невідомих параметрів θ у схемі регресійного експерименту (18) є

$$\hat{\theta} = \arg \min_{\theta \in \Omega} \sum_{j=1}^N [y_j - \eta(x_j, \theta)]^2, \quad (19)$$

4.5.2 Числові методи пошуку МНК-оцінок

Для нелінійної за параметрами функції регресії

$$Q(\theta) = \sum_{i=1}^N [y_i - \eta(x_i, \theta)]^2 \quad (20)$$

система нормальних рівнянь

$$\frac{\partial Q(\theta)}{\partial \theta_i} = 0, \quad i = \overline{1, m}$$

є також нелінійна.

Для пошуку мінімуму функції $Q(\theta)$ використовують як методи пошуку глобального екстремуму (коли немає впевненості у тому, що локальний мінімум у функції Q один), так і стандартні методи пошуку локального екстремуму.

Розглянемо методи, в яких використовуються похідні $\partial \eta(x, \theta) / \partial \theta_i$, ($i = \overline{1, m}$). Принцип побудови цих методів такий самий, як і загальних методів локальної оптимізації. Більша їхня частина записується у вигляді

$$\theta^{(t+1)} = \theta^{(t)} - \gamma_t [F^T(\theta^{(t)})F(\theta^{(t)}) + \alpha_t A]^{-1} F^T(\theta^{(t)})Y(\theta^{(t)}), \quad (21)$$

де $\gamma_t > 0$, $\alpha_t \geq 0$, $\theta^{(t)} \in \Omega$, $t = 0, 1, \dots$, A -- невід'ємно визначена матриця, $\theta^{(0)} \in \Omega$ -- початкове наближення,

$$Y(\theta) = [y_1 - \eta(x_1, \theta), \dots, y_N - \eta(x_N, \theta)]^T,$$

$$F(\theta^{(t)}) = \left\| \frac{\partial \eta(x_i, \theta)}{\partial \theta_j} \Big|_{\theta=\theta^{(t)}} \right\|_{i=\overline{1, N}, j=\overline{1, m}}.$$

Якщо $\alpha_t > 0$, ($t = 0, 1, \dots$), (21) називається *методом Марквардта*, сутність якого у випадку $A = I_m$ полягає у тому, що на кожному кроці методу після лінеаризації (22) і заміни θ_s на $\theta^{(t)}$ замість МНК-оцінки (23) обчислюється гребенева оцінка; якщо $\alpha_t = 0$, $\gamma_t = \arg \min_{\gamma} Q(\theta^{(t+1)})$ -- *методом Хартлі*; якщо $\alpha_t = 0$, $\gamma_t = 1$ -- *методом Гауса--Ньютона*.

Сутність методу Гауса--Ньютона полягає у тому, що функція $\eta(x, \theta)$ апроксимується функцією, лінійною за параметрами в околі точки θ_s

$$\eta(x, \theta) \approx \eta(x, \theta_s) + (\theta - \theta_s)^T \nabla \eta(x, \theta_s). \quad (22)$$

Оцінка МНК для лінеаризованої моделі обчислюється за стандартною формулою, що у даному разі має вигляд

$$\hat{\theta} - \theta_s = [F^T(\theta_s)F(\theta_s)]^{-1} + F^T(\theta_s)Y(\theta_s). \quad (23)$$

Оскільки θ_s -- невідоме, на t -кроці у наведеній формулі θ_s замінюється на $\theta^{(t)}$.

Якщо обчислення (або оцінювання за допомогою скінченних різниць) похідної функції $\eta(x, \theta)$ є трудомістке, економічнішими методами можуть виявитися методи, що не потребують обчислення похідних. Аналог методу Гауса--Ньютона (DUD-метод) має вигляд

$$\theta^{(t+1)} = \theta^{(t)} + \gamma_t (\theta^* - \theta^{(t)}),$$

де

$$\begin{aligned}\theta^* &= m_t \left[\sum_{i=1}^N x_{ti} x_{ti}^T \right]^{-1} \sum_{i=1}^n x_{ti} [y_i - \eta(x_i, \theta^{(t)})], \\ m_t &= \sum_{j=0}^{t-1} \omega_{jt} (\theta^{(j)} - \theta^{(t)}) (\theta^{(j)} - \theta^{(t)})^T, \\ x_{ti} &= \sum_{j=1}^{t-1} \omega_{jt} (\theta^{(j)} - \theta^{(t)}) [\eta(x_i, \theta^{(j)}) - \eta(x_i, \theta^{(t)})].\end{aligned}$$

На практиці вагові множники ω_{jt} , ($t = 0, 1, \dots$; $j = \overline{0, t}$), зазвичай вибираються таким чином:

$$\omega_{jt} = \begin{cases} t^{-1}, & 1 \leq t < m, 1 \leq j < t, \\ m^{-1}, & t \geq m, j \geq t - m - 1, \\ 0, & j < t - m - 1. \end{cases}$$

Сутність DUD-методу полягає у тому, що в околі точки $\theta^{(t)}$ функція $\eta(x, \theta)$ апроксимується лінійною за параметрами:

$$\eta(x, \theta) \approx \eta(x, \theta^{(t)}) + (\theta - \theta^{(t)})^T \gamma_t(x),$$

де

$$\gamma_t(x) = \arg \min_{\gamma > 0} \sum_{j=0}^{t-1} \omega_{jt} [\eta(x, \theta^{(j)}) - \eta(x, \theta^{(t)}) - \gamma(\theta^{(j)} - \theta^{(t)})]^2.$$

Нагадаємо, що у методі Гауса--Ньютона

$$\gamma_t(x) = \frac{\partial \eta(x, \theta)}{\partial \theta} \Big|_{\theta = \theta^{(t)}}.$$

Усі наведені алгоритми мають недолік, властивий більшості загальних методів локального пошуку, і який полягає у тому, що для їхньої збіжності необхідне гарне початкове наближення $\theta^{(0)}$.

Завданням дослідження складних систем і процесів часто є перевірка наявності й встановлення типу зв'язку між незалежними змінними (предикторами, факторами), значення яких можуть змінюватися дослідником і мають деяку заздалегідь задану похибку, та залежною змінною (відгуком). Розв'язування таких задач є предмет регресійного аналізу.

4.5.3 Загальна характеристика методів та задач регресійного аналізу

Класичний регресійний аналіз включає методи побудови математичних моделей систем, що досліджуються, методи визначення параметрів цих моделей і перевірки їхньої адекватності. Він передбачає, що регресія є лінійна комбінація лінійно незалежних базисних функцій від факторів із невідомими коефіцієнтами (параметрами). Фактори й параметри є детерміновані. Відгуки є рівноточні (тобто мають однакові дисперсії) некорельовані випадкові величини. Вважається також, що всі змінні вимірюються у неперервних числових шкалах.

Звичайна процедура класичного регресійного аналізу є така. Спочатку обирають гіпотетичну модель, тобто формулюють гіпотези про фактори, які суттєво впливають на характеристику системи, що досліджується, і тип залежності відгуку від факторів. Потім за наявними емпіричними даними щодо залежності відгуку від факторів оцінюють

параметри обраної моделі. Далі за статистичними критеріями перевіряють її адекватність.

При побудові регресійних моделей реальних систем і процесів вказані вище постулати виконуються не завжди. У більшості випадків їх невиконання призводить до некоректності використання процедур класичного регресійного аналізу та потребує застосування складніших методів аналізу емпіричних даних. Постулат про рівноточність і некорельованість відгуків не є обов'язковий. У разі його невиконання процедура побудови регресійної моделі певною мірою змінюється, але суттєво не ускладнюється.

Складнішою проблемою є обрання моделі та її незалежних змінних. У класичному регресійному аналізі припускають, що набір факторів задається однозначно, всі суттєві змінні присутні в моделі й немає ніяких альтернативних способів обрання факторів. На практиці це припущення не виконується. Тому виникає необхідність розробки формальних і неформальних процедур перетворення та порівняння моделей. Для пошуку оптимальних формальних перетворень використовують методи факторного та дискримінантного аналізу. З розвитком сучасної обчислювальної техніки було розроблено методи послідовної побудови регресійних моделей.

У класичній регресії фактори вважають детермінованими, тобто припускається, що дослідник має про них всю необхідну інформацію із абсолютною точністю. На практиці це припущення не виконується. Відмова від детермінованості незалежних змінних веде до побудови моделей кореляційного аналізу. На жаль, сьогодні їх практичне використання обмежено випадком однофакторних моделей (парною кореляцією). Це пов'язано зі складностями забезпечення та перевірки вимог до багатовимірних функцій розподілу. У деяких випадках можуть використовуватися компромісні методи конфлюентного аналізу, які припускають можливість нормально розподіленого та усікненого розкиду значень факторів. Якщо ця умова виконується, побудову моделі можна звести до багаторазового розв'язування регресійної задачі.

Відмова від припущення про детермінованість параметрів моделей у регресійному аналізі призводить до суттєвих ускладнень, оскільки порушує його статистичні основи. Але на практиці це припущення виконується не завжди. У деяких випадках можливо вважати параметри випадковими величинами із заданим законом розподілу. Тоді як оцінки параметрів можна обрати їх умовні математичні сподівання для відгуків, що спостерігалися. Умовні розподіли та сподівання розраховують за узагальненою формулою Байєса. Тому відповідні методи називають байєсовським регресійним аналізом.

Регресійні моделі часто використовують для опису процесів, що розвиваються в часі. У деяких випадках це призводить до необхідності переходу від випадкових величин відгуків до випадкових послідовностей, випадкових процесів або випадкових полів. Однією з поширених і найпростіших моделей такого типу є модель авторегресії. Вона передбачає, що відгук залежить не тільки від факторів, але також і від часу. Якщо останню залежність можна виявити, то задача зводиться до стандартної задачі побудови регресії для модифікованого відгуку. В інших випадках необхідно використовувати складніші прийоми.

Процедури класичного регресійного аналізу передбачають, що закон розподілу відгуків є нормальний. Проте у практиці найчастішими є випадки, коли цей закон невідомий чи відомо, що він не є нормальний. Їх дослідження зумовило виникнення непараметричного регресійного аналізу, який не передбачає необхідності попереднього завдання функції розподілу.

Важливою проблемою, яка виникає при оцінюванні параметрів регресійних моделей, є наявність грубих помилок у наборі аналізованих даних. Такі помилки можуть виникати внаслідок неправильних дій дослідника, збоїв у роботі апаратури, неконтрольованих короткотривалих сильних зовнішніх впливів на систему тощо.

Використовують два підходи, що дають змогу зменшити вплив грубих помилок на результати аналізу. У першому з них розробляють критерії й алгоритми пошуку помилкових даних. Потім такі дані відкидають. У другому підході розробляють такі алгоритми аналізу, які є нечутливі до наявних помилкових даних (алгоритми робастного оцінювання параметрів). У таких алгоритмах, як критерій оптимальності, часто використовують мінімум суми модулів похибок, або мінімум максимального модуля похибки.

Одним із основних постулатів класичного регресійного аналізу є припущення, що найкращі оцінки параметрів можна отримати, використовуючи метод найменших квадратів. У практиці оцінки, що отримані за допомогою цього методу, часто бувають недостатньо точними і містять великі похибки. Причиною цього може бути структура регресійної моделі. Якщо вона становить собою лінійну комбінацію експонент або є поліном високого степеня, це призводить до поганої зумовленості матриці системи нормальних рівнянь і нестійкості оцінок параметрів. Підвищення стійкості оцінок можна досягти шляхом відмови від вимоги їх незміщеності. Розвиток цього напряму досліджень спричинив виникнення гребеневого, або ріджрегресійного аналізу.

Характерною рисою регресійних моделей, що відрізняє їх від математичних моделей інших типів, є те, що вони не відображають структури систем і процесів, що досліджуються, та присутні в них зв'язки. При побудові регресійних моделей ставиться завдання пошуку статистичних зв'язків між окремими характеристиками системи або між її вхідними та вихідними параметрами.

Найчастіше задача побудови регресійної моделі формулюється так. Необхідно знайти функцію заданого класу, для якої функціонал

$$F(\alpha) = \sum_{i=1}^n [z_i(\alpha, X) - y_i]^2 \rightarrow \min. \quad (24)$$

У цьому виразі $z_i(\alpha, X)$ -- значення функції, що апроксимує залежність в i -й точці, y_i -- відповідне значення емпіричної залежності, α -- вектор параметрів, які треба знайти, X -- вектор незалежних змінних. Отриману функцію називають регресійною моделлю. Метод її пошуку, що ґрунтується на використанні критерію (24), називають методом найменших квадратів.

Функцію, що апроксимує, у разі однієї незалежної змінної часто шукають у вигляді поліному $z(x) = \sum_{j=1}^M \alpha_j x^j$, зворотного поліному $z(x) = 1/\sum_{j=1}^M \alpha_j x^j$, експоненційних або показникових функцій $z = ae^x$ чи $z = ab^x$, степеневій функції $z = \alpha x^b$, лінійно-логарифмічній функції $z = \alpha_1 + \alpha_2 x + \alpha_3 \ln x$, тригонометричного ряду Фур'є тощо. У разі кількох незалежних змінних найчастіше використовують моделі лінійні як за параметрами, так і за незалежними змінними $z(x) = \alpha_0 + \sum_{j=1}^p \alpha_j x^j$, а також поліноміальні моделі, що лінійні за параметрами, але нелінійні за незалежними змінними.

$$z = \alpha_0 + \sum_{i,j}^p \alpha_{i,j} x_i x_j + \sum_{i,j,k}^p \alpha_{i,j,k} x_i x_j x_k + \dots$$

Останні відповідають розкладенню функції відгуку у ряд Тейлора. Проте можливе й використання для апроксимації інших видів залежностей.

Регресійні моделі називають лінійними або нелінійними, якщо вони є лінійні або нелінійні за параметрами. При цьому визначення "лінійна" часто опускають. Значення

найвищого степеню предиктора в поліноміальних моделях називають порядком моделі. Наприклад,

$$z = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \varepsilon,$$

де ε -- похибка моделі, є лінійна модель третього порядку.

Обрання функції, що апроксимує, є нетривіальна задача. Спочатку рекомендують нанести наявні емпіричні дані на графік. Це дає змогу визначити наявність чи відсутність залежності між величинами, що досліджуються, а також зробити деякі припущення про можливий тип такої залежності. Після того, як було встановлено наявність кореляції між досліджуваними величинами, виникає необхідність підбору функції, що апроксимує одержану залежність.

Важливою особливістю регресійних моделей є те, що їх не можна використовувати за межами тієї області значень вихідних параметрів, для якої була побудована ця модель. При побудові регресійних моделей типу поліному, зворотного поліному, тригонометричного ряду та деяких інших слід мати на увазі, що, збільшуючи кількість членів ряду, можна отримати скільки завгодно близьке до нуля значення функціоналу (18). Проте це не завжди свідчить про якість апроксимації, оскільки цей функціонал не дає інформації про ступінь близькості апроксимуючої функції та емпіричної залежності у проміжках між наявними точками. Найпростішим критерієм, що дозволяє обмежити кількість членів ряду, який апроксимує емпіричну залежність, є порівняння відношення значення функціоналу (18) до суми квадратів похибок емпіричних значень функції, а також величини, зворотної до цього відношення, із значенням критерію Фішера для відповідного числа ступенів вільності. Якщо перше відношення перевищує значення критерію Фішера, це свідчить про недостатню кількість членів ряду. Якщо значення критерію Фішера є менше за друге відношення, можна зробити висновок про надмірно велику кількість членів ряду. Іншою проблемою може бути наявність кількох екстремумів функціонала (18). У цьому разі необхідно мати на увазі, що більшість стандартних алгоритмів дозволяють знаходити локальні, а не глобальні екстремуми функціоналів. Тому результат аналізу залежить від вибору початкових умов.

Регресійні моделі, що побудовані на базі поліномів, як правило, є формальні. Їх використовують для опису систем і процесів, теорію яких розроблено недостатньо. Цікавішими для дослідників зазвичай є змістовні моделі, що відображають структуру та зв'язки у системах, сутність та механізми процесів. Якщо теоретичні основи досліджуваних систем і пошуку процесів достатньо розроблені, експериментальні дані потрібні для визначення окремих параметрів моделі.

Як правило, змістовні моделі бувають нелінійними за параметрами. Методологію їх дослідження розробляє нелінійний регресійний аналіз. Для задач, що розв'язуються в межах цього напрямку, характерними є також й інші ускладнення. Зокрема, моделі зазвичай бувають багатовимірними. При цьому окремі відгуки можуть бути зв'язані один з одним. Сама регресійна модель часто задається у неявному вигляді і є неаналітичний розв'язок системи алгебраїчних або диференціальних рівнянь. Нестійкість оцінок параметрів у нелінійних моделях різко зростає. Як правило, такі задачі мають кілька розв'язків або не мають розв'язків взагалі.

Аналогічні ускладнення можуть виникнути при побудові функціональних моделей, коефіцієнти яких необхідно визначати шляхом апроксимації емпіричних даних. Але у цьому разі потрібно застосовувати інші критерії адекватності моделі. Важливою умовою адекватності моделі є відповідність загальної поведінки моделі, що апроксимує, реальній поведінці об'єкту дослідження у подібних умовах. Крім того, необхідно, щоб за малих змін вихідних даних параметри моделі, що підбираються при апроксимації, змінювалися

мало або залишалися незмінними. Ще однією складністю, що виникає при аналізі таких моделей, є можливість отримання значень параметрів, які є неприпустимі з точки зору змістовної постановки задачі. Наприклад, може бути отримано від'ємні значення параметрів, що мають лише невід'ємні значення. Тому задачу пошуку екстремуму функціоналу (73) у цьому разі треба ставити як задачу умовної оптимізації, а не безумовної, як при побудові регресійних моделей.

Регресійний аналіз застосовується для побудови моделей зв'язку вихідних величин процесу із вхідними величинами, що є випадкові, з метою передбачення поведінки вихідних величин процесу залежно від конкретних значень вхідних величин.

Нехай модель процесу представляється рівнянням виду

$$y = f(x)\alpha, \quad (25)$$

де $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_m]^T$ -- вектор невідомих параметрів моделі, $x = [x_1, x_2, \dots, x_n]^T$ -- вектор значень незалежних змінних, $f = [f_0(x), f_1(x), \dots, f_n(x)]$ -- вектор заданих функцій.

За результатами експерименту необхідно знайти оцінку a параметрів α . Тоді оцінка $\hat{y}$ для функції y обчислюється за формулою

$$\hat{y} = f(x)a, \quad (26)$$

Нехай експеримент проводиться у N точках $x^1, x^2, \dots, x^N$. Позначимо y вектор, координатами якого є результати його спостережень у точках x^k :

$$y = [\tilde{y}^1, \tilde{y}^2, \dots, \tilde{y}^N]^T.$$

Власне кажучи, у кожній точці x^k може бути виконано ν експериментів, результатами яких будуть спостереження

$$\tilde{y}^{k_1}, \tilde{y}^{k_2}, \dots, \tilde{y}^{k_\nu}.$$

Тоді

$$y^k = \frac{1}{\nu} [\tilde{y}^{k_1}, \tilde{y}^{k_2}, \dots, \tilde{y}^{k_\nu}].$$

Задача -- на підставі результатів спостережень y^k знайти найкращі у деякому сенсі оцінки a та y .

Теоретичні основи методу.

За основу методу побудови моделі покладено метод найменших квадратів. Позначимо

$$y^k = \frac{1}{\nu} [\tilde{y}^{k_1}, \tilde{y}^{k_2}, \dots, \tilde{y}^{k_\nu}]^T.$$

-- вектор значень, який обчислено у відповідності з ((136)).

$$\hat{y}^k = f^T a, \quad k = \overline{1, N}, \quad (27)$$

або у матричному вигляді

$$\hat{y} = Fa, \quad (28)$$

$$F = \begin{bmatrix} f_0(x^1) & f_1(x^1) & \dots & f_m(x^1) \\ f_0(x^2) & f_1(x^2) & \dots & f_m(x^2) \\ \dots & \dots & \dots & \dots \\ f_0(x^N) & f_1(x^N) & \dots & f_m(x^N) \end{bmatrix}.$$

Будемо вважати, що результати експерименту y^k розподілені нормально, є

випадкові величини, що досліджуються, тобто математичне сподівання величини y^k відповідає дійсному значенню

$$M[y] = y = Fa.$$

Нехай також результат спостережень у точці x^k не залежить від результату спостережень у точці x^j й дисперсія результатів спостережень y^k однакова у всіх точках x^k . Останні дві умови виконуються, якщо

$$M[(y - y^k)(y - y^k)^T] = \sigma^2 E.$$

Крім того, припустимо, що оцінка a , що являє собою випадковий вектор, є незміщена, тобто $M[a] = a$, а дисперсія σ_k^2 оцінки $\hat{a}_k$ має бути мінімальною:

$$\sigma_k^2 = M[(\hat{a}_k - a_k)^2], \quad k = \overline{0, n}.$$

Оцінка a задовольняє цим умовам, якщо

$$S = (y - y^k)^2 = |y^k - y|^2 \quad (29)$$

є мінімальна.

З урахуванням (27) маємо

$$S = S(bfa) = y^T y + a^T F^T F a - 2y^T F a.$$

Якщо матриця $F^T F$ не вироджена, тобто якщо матриця F має ранг $m+1$, сума квадратів (5) має один мінімум при

$$a = CF^T y. \quad (30)$$

Матриця $C = [F^T F]^{-1}$ має назву дисперсійної.

Статистичний аналіз точності.

Похибка оцінки a , що обчислена згідно з (84) за методом найменших квадратів, тим більше, чим більше дисперсія похибок спостережень. Точність оцінки $\hat{a}_k$ та величини y характеризуються відповідно дисперсіями $\sigma_{a_k}^2$ та σ_y^2 , що залежать від матриці F . З метою визначення цієї залежності необхідно знайти вираз для коваріаційної матриці $V(a)$:

$$V(a) = M[(a - a)(a - a)^T] = C\sigma^2.$$

Звідси для дисперсії $\sigma_{a_k}^2$ оцінки $\hat{a}_k$ маємо

$$\sigma_{a_k}^2 = c_{kk} \sigma^2, \quad \sigma_y^2 = \sum_{k=1}^N (x) C f(x) \sigma^2.$$

Оцінка y визначається за формулою (85).

Далі, згідно з попереднім розділом, визначаються довірчі інтервали для оцінок параметрів a .

Перевірка адекватності моделі.

Розглянемо величину

$$S_3 = \sum_{k=1}^N (y^k - y)^2 = \nu S_1,$$

що має $K_3 = N - m - 1$ ступенів свободи.

Частка від ділення оцінки дисперсії неадекватності на оцінку дисперсії похибки одиничного спостереження

$$F = \frac{S_3/K_3}{S_2/K_2} \quad (31)$$

у випадку, коли модель адекватна, є випадкова величина, що задовольняє $\mathcal{F}$ - розподілу Фішера із числом ступенів свободи K_3 та K_2 . Для визначеного рівня значимості α перевірки гіпотези про адекватність моделі можна визначити значення $\mathcal{F}_{kr}$, що відповідає умові

$$P\{\mathcal{F} > \mathcal{F}_{kr}\} = \alpha = 1 - p. \quad (32)$$

Реалізація побудови моделі.

- Вибір моделі.

$$y(x) = a_0 F_0(x) + a_1 f_1(x) + \dots + a_m f_m(x).$$

- Для $N > m + 1$ точок

$$x^k = [x_1^k, x_2^k, \dots, x_n^k]^T, \quad k = \overline{1, N}$$

обчислюється матриця (28).

- Обчислюється дисперсійна матриця $C = [F^T F]^{-1}$.

- У кожній точці x^K здійснюються ν випробувань, за результатами яких обчислюється середнє значення

$$y^k = \frac{1}{\nu} \sum_{j=1}^{\nu} \tilde{y}^{kj}$$

із дисперсією σ^2 .

- Обчислюються оцінки a коефіцієнтів моделі за методом найменших квадратів

$$a = CF^T y.$$

- Обчислюються дисперсії оцінок $\sigma_k = c_{kk} \sigma^2$.
- Визначається оцінка функції $\hat{y}$ за формулою ((137)).
- Обчислюється оцінка дисперсії похибок спостережень s :

$$s^2 = S_2 / (\nu K_2), \quad K_2 = N(\nu - 1),$$

$$S_2 = \sum_{k=1}^N \sum_{j=1}^N (\tilde{y}^{kj} - \tilde{y}^k)^2, \quad s^2 + k = c_{kk} \sigma^2.$$

- Обчислюється

$$S_3 = \sum_{k=1}^N \nu (\tilde{y}^k - \hat{y}^k)^2.$$

- Оцінюється адекватність моделі

$$\mathcal{F} = \frac{S_3 / K_3}{S_2 / K_2}, \quad K_3 = N - m - 1.$$

- Якщо дисперсія спостережень σ^2 відома й модель адекватна, довірчий інтервал із довірчою імовірністю p визначається за нерівністю

$$|\hat{a}_k - a_k| < \sigma_k^2 \varepsilon.$$

Значення ε визначаємо із табл. В.1 за умови $\Phi(\varepsilon) = p/2$. Якщо дисперсія спостережень σ^2 не відома й у кожній точці плану проводиться один експеримент,

$$|\hat{a}_k - a_k| < s_k \varepsilon, \quad s_k^2 = c_{kk} s^2, \quad k = \overline{0, m},$$

$$\bar{s}^2 = S_1 / K_1 = \sum (\tilde{y}^k - \hat{y}^k)^2 / K_1,$$

ε знаходимо із табл. В.5 при $K_1 = N - m - 1$ ступенях свободи й заданій довірчій імовірності p .

- Перевіряються коефіцієнти a_k на значимість. Якщо

$$|\hat{a}_k| > t_{kr} s_k,$$

коефіцієнт a_k вважається значимим; t_{kr} -- критичне значення розподілу Ст'юдента для заданого рівня значимості α та K_1 ступенів свободи. t_{kr} визначається із табл. В.4 за умови $P\{|t| > t_{kr}\} = \alpha$. Коефіцієнти, які виявились незначимими, відкидаються.

4.6 Плани для квадратичних моделей

Побудова лінійних відносно факторів регресійних моделей має поширене застосування при аналізі, наприклад, моделей в економіці, з огляду на простоту їх реалізації, але які зовсім не задовольняють умовам адекватності моделей реальним процесам.

Розглянемо плани для квадратичних регресійних моделей

$$y(a; x) = a_0 + \sum_{i=1}^n [a_i x_i + a_{i+n} x_i^2] + a_{2n+1} x_1 x_2 + \dots + a_K x_{n-1} x_n, \quad (33)$$

$$(K+1) = \frac{1}{2}(n+2)(n+1).$$

Для оцінки коефіцієнтів моделі незалежна змінна у плані має набувати три значення. Композиційний план для квадратичних моделей можна отримати шляхом додавання деякої кількості спеціальних точок до ядра, утвореного планом для лінійної моделі. Якщо до ядра додати точку у центрі плану із координатами $0, \dots, 0$ та $2n$ так званих "зіркових точок" із координатами

$$(\pm\alpha, 0, \dots, 0), \dots, (0, \dots, 0, \pm\alpha),$$

отримаємо центральний композиційний план, запропонований Боксом.

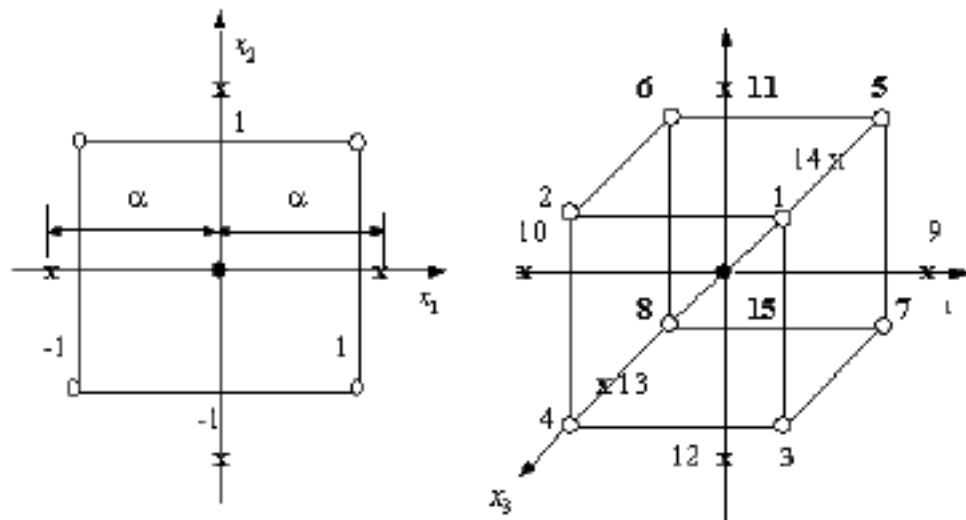


Рис. 4.1. Композиційний план

Для забезпечення ортогональності плану необхідно перетворити модель (33)

$$y(a; x) = a_0 + \sum_{i=1}^n [a_i x_i + a_{i+n} (x_i^2 - \beta)] + a_{2n+1} x_1 x_2 + a_K x_{n-1} x_n, \quad (34)$$

$$\beta = \frac{1}{N} \sum_{j=1}^N (x_i^j)^2 = \frac{1}{N} [2^{n-p} + \alpha^2]. \quad (35)$$

У цьому виразі N -- загальна кількість точок у плані; 2^{n-p} -- число точок ядра плану. Якщо покласти

$$a_0 = b_0 - \beta \sum_{j=1}^n a_{n+j}, \quad (36)$$

можна від моделі (34) перейти до моделі (33). Загальна кількість точок плану

$$N = 2^{n-p} + 2n + 1.$$

У якості ядра у композиційних планах можуть бути використані тільки такі дробові плани, що дозволяють отримати оцінки коефіцієнтів при парних незмішаних взаємодіях. Оскільки вибір величини α не впливає на ортогональність стовпців матриці F , що відповідають факторам x_i та парним добуткам $x_i x_j$, для ортогональності композиційного плану необхідно вимагати, щоб парні добутки у ядрі плану не були змішані із лінійними членами. Неважко переконатися, що для $n \leq 4$ дробових планів, які задовольняють зазначеним вимогам, не існує.

Для $n = 5$ як ядро можна використати дробовий факторний план 2^{5-1} із генератором $x_5 = x_1 x_2 x_3 x_4$; для $n = 6$ або $n = 7$ також можна використати тільки план 2^{n-1} . Тільки для $n = 8$ можливе застосування плану 2^{8-2} , наприклад, із генераторами $x_7 = x_1 x_2 x_3 x_4$ та $x_8 = x_1 x_2 x_5 x_6$.

Перейдемо тепер до вибору плеча ортогональних планів. Скалярні добутки будь-яких двох стовпців матриці F дорівнюють нулю за будь-якого вибору α . Виняток складають стовпці при квадратичних функціях вигляду $(x_i^2 - \beta)$; їхні парні добутки у загальному випадку не дорівнюють нулю, а залежать від α .

Прирівнявши до нуля суму добутків елементів двох стовпців при функціях вигляду $(x_i^2 - \beta)$, отримаємо умову для вибору значення α , що забезпечує ортогональність плану:

$$\begin{aligned} & 2^{n-p} \left(1 - \frac{1}{N} [2^{n-p} + 2\alpha^2] \right)^2 - \frac{4}{N} [2^{n-p} + 2\alpha^2] \left(\alpha^2 - \frac{1}{N} [2^{n-p} + 2\alpha^2] \right) \\ & + (2n-3) \left(\frac{1}{N} [2^{n-p} + 2\alpha^2] \right)^2 = 0. \end{aligned} \quad (37)$$

Звідси випливає

$$\alpha = \sqrt{2^{(n-p)/2-1} (\sqrt{N} - 2^{(n-p)/2})}. \quad (38)$$

Інформаційна матриця плану:

$$M = \begin{bmatrix} m_0 & 0 & 0 & 0 \\ 0 & m_1 E_n & 0 & 0 \\ 0 & 0 & m_2 E_n & 0 \\ 0 & 0 & 0 & m_3 E_{\binom{n}{2}} \end{bmatrix}. \quad (39)$$

$$m_0 = N = 2^{n-p} + 2n + 1, \quad m_1 = 2^{n-p} 2\alpha^2,$$

$$m_2 = 2^{n-p} (1 - \beta)^2 + 2(\alpha^2 - \beta)^2 + (2n-1)\beta^2, \quad m_3 = 2^{n-p}. \quad (40)$$

Звідси отримуємо вираз для дисперсійної матриці:

$$C = \begin{bmatrix} c_0 & 0 & 0 & 0 \\ 0 & c_1 E_n & 0 & 0 \\ 0 & 0 & c_2 E_n & 0 \\ 0 & 0 & 0 & c_3 E_{\binom{n}{2}} \end{bmatrix}. \quad (41)$$

де $c_i = 1/m_i$.

Табл. 4.1. Параметри ортогональних центральних композиційних планів

n	Ядро плану	N	α	β	c_0	c_1	c_2	c_3
2	2^2	9	1	2/3	0,1111	0,1667	0,5	0,25
3	2^3	15	1,215	0,73	0,0667	0,0913	0,2298	0,1250
4	2^4	25	1,414	0,8	0,040	0,050	0,125	0,0625
5	2^{5-1}	27	1,547	0,77	0,03704	0,0481	0,0871	0,0625
6	2^{6-1}	45	1,722	0,843	0,0222	0,0264	0,0564	0,03125
7	2^{7-1}	79	1,885	0,900	0,0127	0,0141	0,0389	0,0158
8	2^{8-2}	81	2,001	0,8889	0,0123	0,0139	0,0312	0,0156

Формули для оцінки коефіцієнтів регресійної моделі:

$$\hat{a}_i = \begin{cases} c_1 \sum_{j=1}^N x_i^j \tilde{y}^j, & i = \overline{1, n} \\ c_2 \sum_{j=1}^N [(x_i^j)^2 - \beta] \tilde{y}^j, & i = \overline{n+1, 2n} \\ c_3 \sum_{j=1}^N x_\mu^j x_\lambda^j \tilde{y}^j, & \mu, \lambda = \overline{1, n}, \mu \neq \lambda, i = \overline{2n+1, K}. \end{cases} \quad (42)$$

Оцінка $\hat{b}_0$ обчислюється за формулою:

$$\hat{b}_0 = \frac{1}{N} \sum_{j=1}^N \tilde{y}^j. \quad (43)$$

$$\hat{a}_0 = \hat{b}_0 - \beta \sum_{j=1}^n \hat{a}_{n+j}. \quad (44)$$

Оцінка дисперсій коефіцієнтів визначається за формулою:

$$s_i^2 = \begin{cases} s^2 c_0, & i = 0, \\ s^2 c_1, & i = \overline{1, n} \\ s^2 c_2, & i = \overline{n+1, 2n} \\ s^2 c_3, & i = \overline{2n+1, K}. \end{cases} \quad (45)$$

s -- дисперсія спостережень. Для $\hat{a}_0$ маємо

$$s_{\hat{a}_0}^2 = s^2 (c_0 + n\beta^2 c_2). \quad (46)$$

Приклад. Виконується повний факторний експеримент для встановлення залежності відгуку від факторів, що наведені в табл. 4.2. У загальному вигляді нормування фактору виражається формулою

$$\tilde{x} = \frac{2(x_i - x_{ic})}{x_{ih} - x_{il}} = \frac{x_i - x_{ic}}{\Delta}, \quad \Delta = \frac{1}{2}(x_{ih} - x_{il}).$$

x_{ih} -- верхнє значення параметру x , x_{il} -- нижнє значення параметру x .

Табл. 4.2.

X_1	60	40	0	0	0	0	0	0
X_2	0	0	80	20	0	0	0	0
X_3	0	0	0	0	10	0	10	0
Y	2	3	4	5	4	5	8	7

Проведено N дослідів. Маємо ПФЕ 2^3 . Визначимо центр плану та інтервали варіювання факторів.

$$X_1^{\max} = 60; X_1^{\min} = 40; X_1^0 = \frac{60+40}{2} = 50; \Delta X_1^0 = \frac{60-40}{2} = 10;$$

$$X_2^{\max} = 80; X_2^{\min} = 20; X_2^0 = \frac{80+20}{2} = 50; \Delta X_2^0 = \frac{80-20}{2} = 30;$$

$$X_3^{\max} = 10; X_3^{\min} = 0; X_3^0 = \frac{10+0}{2} = 5; \Delta X_3^0 = \frac{10-0}{2} = 5.$$

Переходимо до кодованих змінних:

$$x_1 = \frac{X_1 - 50}{10}, x_2 = \frac{X_2 - 50}{30}, x_3 = \frac{X_3 - 5}{5}.$$

Запишемо результати експерименту у табл. 4.3.

Табл.4.3 ПФЕ

№	x_0	x_1	x_2	x_3	y
1	+	+	-	-	2
2	+	-	-	-	3
3	+	+	+	-	4
4	+	-	+	-	5
5	+	+	-	-	4
6	+	-	-	+	5
7	+	+	+	+	8
8	+	-	+	+	7

$$b_k = \frac{1}{2^n} \sum_{i=1}^{2^n} y_k \tilde{x}_{k,i}.$$

$$\hat{y} = \sum_{k=0}^{m=3} b_k x_k, (x_0 = 1).$$

Для прикладу, що розглядається,

$$\hat{y} = 4,75 - 0,25x_1 + 1,25x_2 + 1,25x_3.$$

У натуральних змінних

$$\hat{y} = 4,75 - 0,25 \frac{X_1 - 50}{10} + 1,25 \frac{X_2 - 50}{30} + 1,25 \frac{X_3 - 5}{5}.$$

Запишемо результати експерименту у табл. 4.4.

Табл. 4.4. Матриця ПФЕ

№	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	y
1	+	+	-	-	-	-	+	2
2	+	-	-	-	+	-	+	3
3	+	+	+	-	+	-	-	4
4	+	-	+	-	+	-	-	5
5	+	+	-	+		-	-	4
6	+	-	-	+	+	-	-	5
7	+	+	+	+	+	+	+	8
8	+	-	+	+	-	-	+	7
b_j	4,75	-0,25	1,25	0,25	0,25	0,25	0,25	

$$\hat{y} = 4,75 - 0,25x_1 + 1,25x_2 + 1,25x_3 + 0,25x_1x_2 + 0,25x_1x_3 + 0,25x_2x_3.$$

4.7 Композиційні ортогональні плани

4.7.1 Повний факторний експеримент (ПФЕ)

При статистичному методі планування експерименту існує правило -- число рівнів варіювання факторів має бути як мінімум на одиницю більше порядку поліному.

Проведення експерименту має забезпечити мінімальний вплив випадкових параметрів процесу, що досліджується, на функцію відгуку. Для цього треба дотримуватися таких вимог:

-- передбачити проведення кількох паралельних дослідів за одних і тих самих умов, що передбачаються відповідним рядком матриці планування (номером досліду);

-- необхідно рандомізувати параметри, що не контролюються, процесу, тобто забезпечити їхню взаємну компенсацію.

Результати n паралельних дослідів усереднюють і при аналізі результатів експерименту використовують усереднене значення функції відгуку:

$$\bar{Y}_j = \frac{1}{n} \sum_{i=1}^n Y_{ji}, \quad j = \overline{1, N}.$$

Для виконання другої вимоги порядок реалізації умов досліду (перший стовпець матриці) має бути рандомізований. Для цього для кожної із n серій дослідів за допомогою таблиці випадкових чисел визначається послідовність дослідів.

Обробка і аналіз результатів ПФЕ передбачає такий порядок їхнього проведення:

1. Оцінюються дисперсії середнього арифметичного у кожному рядку матриці:

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_{ji} - \bar{Y}_j)^2.$$

2. Перевіряються однорідності дисперсій. Здійснюється за критерієм Кохрена. Підраховують параметр

$$G = \frac{\max_j S_j^2}{\sum_{j=1}^N S_j^2},$$

тобто обчислюють відношення максимального значення мінливості (максимального

значення дисперсії поміж N дослідів до суми мінливостей в усіх дослідах). Це значення G порівнюють із критичним (табличним) його значенням G_{kr} .

Якщо $G \leq G_{kr}$, G є результат випадкового розсіювання функції відгуку, тобто експерименти відновлювані й їхні результати можна використовувати для оцінки коефіцієнтів регресії. Якщо $G > G_{kr}$, експерименти не відновлюються, тобто неконтрольовані і некеровані фактори створюють на виході занадто великий рівень "шуму".

3. Створюється модель об'єкту із перевіркою статистичної значимості коефіцієнтів поліному.

Після виконання ПФЕ здійснюють незалежну оцінку коефіцієнтів поліному за формулою:

$$b_i = \frac{1}{N} \sum_{j=1}^N X_{ij} \bar{Y}_j,$$

X_{ij} набуває значень ± 1 відповідно до матриці планування.

Після обчислення коефіцієнтів оцінюється їх значимість для визначення ступеню впливу різних факторів на вихідний параметр (функцію відгуку). Підґрунтям оцінки значимості є співставлення абсолютного значення коефіцієнтів, наприклад, b_i і дисперсії похибки його визначення $S^2\{b_i\}$. У цьому разі за допомогою t -критерію (критерію Ст'юдента) перевіряється гіпотеза про незначимість коефіцієнту (тобто гіпотези про $b_i = 0$ (перевірка нуль-гіпотези) за формулою

$$t_i = |b_i| / \sqrt{s^2(b)}.$$

Коефіцієнт b_i вважається значимим, коли $t \geq t_{kr}$ для числа ступенів свободи $\nu = N(n-1)$.

4. Перевіряється адекватність. Для цього досить оцінити відхилення значення вихідного параметра Y_{jt} від результатів експерименту $\bar{Y}_j$ у точці X_j факторного простору.

Оцінюємо дисперсію адекватності за формулою:

$$S_{ad}^2 = \frac{1}{N-d} \sum_{j=1}^N (\bar{Y}_j - Y_{jt})^2,$$

d -- кількість членів поліному, що апроксимує.

Якщо $S_{ad}^2 \leq S^2\{Y\}$, отримана модель адекватна результатам експерименту, якщо $S_{ad}^2 > S^2\{Y\}$, перевірка гіпотези про адекватність виконується за допомогою F -критерію (Фішера) при $\nu_{ad} = N - d$ $\nu = N(n-1)$.

$$F = S_{ad}^2 / S^2\{Y\},$$

Якщо $F \leq F_{kr}$, модель вважається адекватною процесу.

4.7.2 ЦКП

З теорії інтерполяції відомо, що для відшукування коефіцієнтів, що розділені, інтерполяційного поліному як імітаційної моделі процесу число рівнів змінювання кожної із незалежних змінних має бути на одиницю більше за порядок поліному. У ПФЕ при

$k = 2$ їх має бути $3^2 = 9$, а для 3-х факторів $3^3 = 27$. Тому за збільшення числа факторів, що враховуються, застосування ПФЕ 3^k призводить до різкого збільшення обсягу експерименту.

Скоротити число дослідів можна за використання центральних композиційних планів (ЦКП), ядром яких є лінійні ортогональні плани. Перевага цих планів полягає у тому, що якщо гіпотеза про лінійність моделі, що відповідає процесу, у результаті аналізу експериментальних даних не підтверджується, нема необхідності ставити усі експерименти заново для отримання моделі вищого порядку, а досить додати кілька спеціально спланованих експериментальних точок так, щоб отримати план, що відповідає поліному другого порядку.

Тоді у центрі плану, що відповідає початковому значенню усіх факторів, що враховуються в експерименті, проводиться дослід, умови якого у матриці планування експерименту відображаються нулями для безрозмірних величин усіх факторів.

Отже, у ПФЕ 2^3 до 8-ми дослідів додаються ще 7 дослідів (включаючи дослід у центрі плану), 6 із яких відповідають "зірковим точкам". Ці точки являють собою два рівня варіювання кожним із 3-факторів, значення яких містяться за межами граней кубу і належать поверхні сфери радіусу 2а.

Загальна кількість дослідів ЦКП при k факторах складає

$$N = 2^k + 2k + m_0,$$

де $2k$ -- кількість "зіркових точок"; m_0 -- число дослідів у центрі плану, а загальна кількість рівнів варіювання ЦКП дорівнює 5.

Найпоширеніший тип ЦКП -- *центральний композиційний ротатабельний план (ЦКРП)*. При цьому

$$\sigma^2[Y] \rightarrow \text{const} \quad \text{çà} R = \text{const}.$$

ЦКРП надає можливість мінімізувати похибки у визначенні Y із-за неадекватності подання результатів дослідження процесу імітаційною моделлю у вигляді поліному 2-го порядку. Це досягається тим, що при виборі "зіркових точок" на осях координат для неперервності інформаційної поверхні вони доповнюються інформацією у центрі плану, що є сфера із нульовим радіусом, тобто інформацією, рівноточною в усіх напрямках. Питома вага цієї інформації у загальному обсязі інформації збільшується, що досягається збільшенням числа дослідів (τ_0) у центрі плану. При цьому інформаційні поверхні наближаються до сфер. Для $k = 3$ $\tau_0 = 6$ (тобто числу "зіркових точок"). Це призводить до збільшення числа дослідів, але забезпечує неперервність інформаційної поверхні та її ідентичність незалежно від повороту осей координат.

Кількість дослідів у центрі плану залежить від кількості факторів. При $k = 3$ $N - 0 = 6$ збігається із числом зіркових точок. Це призводить до збільшення кількості дослідів порівняно із ЦКОП, але забезпечує неперервність інформаційної поверхні та її ідентичність незалежно від повороту осей координат.

При реалізації ротатабельних планів можна відмовитися від постановки паралельних дослідів для оцінки відновлюваності експериментів, що зменшує загальну кількість дослідів. Дисперсія відновлюваності (дисперсія експериментальних значень функції відгуку у паралельних дослідах) оцінюється у цьому разі за експериментами у центрі плану.

Табл 4.5. Параметри ротатабельних центральних композиційних планів

n	Ядро плану	n_z	n_0	N	α
2	2^2	4	5	13	1,414
3	2^3	6	6	20	1,682
4	2^4	8	7	31	2,000
5	2^5	10	10	52	2,378
	2^{5-1}	10	6	32	2,000
6	2^6	12	15	91	2,828
	2^{6-1}	12	9	52	2,378
7	2^7	14	21	163	3,333
	2^{7-1}	14	14	92	2,828

Формули для розрахунку коефіцієнтів поліному та їхні дисперсії при ротатабельному плануванні:

$$b_0 = \frac{A}{N} \left[2\lambda^2 \left(k + 2(0y) - 2\lambda c \sum_{i=1}^k (iiy) \right) \right];$$

$$b_i = \frac{c}{N} (iy); \quad b_{ij} = \frac{c^2}{N\lambda} (ijy);$$

$$b_{ii} = \frac{A}{N} \left[c^2 [(k+2)\lambda - k](ij) + c^2 (1-\lambda) \sum_{i=1}^k (ii) - 2\lambda c (0y) \right];$$

$$\sigma^2[b_0] = 2A\lambda^2(k+2) \frac{\sigma^2[Y]}{Nn}; \quad \sigma^2[b_i] = \frac{c\sigma^2[Y]}{Nn}; \quad \sigma^2[b_{ij}] = \frac{c^2\sigma^2[Y]}{\lambda Nn}.$$

У цих формулах позначено:

$$c = N / \sum_{j=1}^N x_{ij}^2; \quad A = \frac{1}{2\lambda[(k+2)\lambda - k]}; \quad \lambda = \frac{kN}{(k+2)n_1}; \quad n_1 = N - m_0;$$

$$(0y) = \sum_{j=1}^N x_{0j} y_j; \quad (iiy) = \sum_{j=1}^N x_{ij}^2 y_j;$$

$$(iy) = \sum_{j=1}^N x_{ij} y_j; \quad (ijy) = \sum_{l=1}^N x_{il} x_{jl} y_l.$$

4.7.3 Ротатабельні центральні композиційні плани

Ротатабельний центральний композиційний план будується аналогічно ортогональному плану. За ядро можна використати повний або дробовий факторний план. Величина плеча α визначається за формулою

$$\alpha = 2^{(n-p)/2}. \quad (47)$$

Кількість спроб n_0 у центрі плану вибирається із таких міркувань. Висувається вимога незмінності інформації про значення вихідної змінної для точок плану в середині сфери одиничного радіусу із центром у центрі плану. Плани, що задовольняють цю вимогу, називаються *ротатабельними уніформ-планами*.

Виявляється, що уніформ-план можна отримати, змінюючи число точок у центрі ротатабельного плану. В табл. 4.5 наведено значення плеча α та числа n_0 точок у центрі плану, зіркових точок n_z та загального числа точок N для ротатабельних уніформ-планів другого порядку.

Зазначимо, що у матриці F центральна точка фігурує n_0 разів. Це означає, що при обчисленні оцінок коефіцієнтів використовується результат кожного виміру у центрі плану, а не їхнє середнє значення. Така побудова матриці F є наслідком вимоги однакової дисперсії всіх величин залежної змінної, що входять у матрицю $\tilde{Y}$. Паралельні спроби у центрі плану дозволяють розраховувати оцінку дисперсії похибок спостереження.

У загальному випадку за наявності n_0 точок у центрі плану і повторенні експерименту ν разів у кожній точці матриці F , оцінки дисперсії одиничного експерименту s_e^2 визначаються за формулою

$$s_e^2 = S_e / \varphi_e. \quad (48)$$

$$S_e = \sum_{j=1}^{N1} \sum_{i=1}^{\nu} (\tilde{y}^j - \tilde{y}^{ji})^2 + \sum_{j=1}^{n_0} \sum_{i=1}^{\nu} (\tilde{y}^0 - \tilde{y}^{N1+j,i})^2. \quad (49)$$

$$N1 = 2^{n-p} + 2n, \quad \tilde{y}^0 = \frac{1}{n_0 \nu} \sum_{j=1}^{n_0} \sum_{i=1}^{\nu} \tilde{y}^{N1+j,i}. \quad (50)$$

$$\varphi_e = N1(\nu - 1) + n_0 \nu - 1. \quad (51)$$

$$y_e = [8, 27.9, 5.8, 43, 14.1, 7, 20.2, 30.2, 12.1, 36, 10.4, 25.3, 20, 18, 29, 28.4, 28.6, 28.8].$$

Табл. 4.6. Ротатабельний центральний композиційний план та матриця F , $n = 3$

x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$
1	-1	-1	-1	1	1	1	1	1	1
1	1	-1	-1	1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1	1	-1
1	1	1	-1	1	1	1	1	-1	-1
1	-1	-1	1	1	1	1	1	-1	-1
1	1	-1	1	1	1	1	-1	1	-1
1	-1	1	1	1	1	1	-1	-1	1
1	1	1	1	1	1	1	1	1	1
1	+1,682	0	0	2,828	0	0	0	0	0
1	-1,682	0	0	2,828	0	0	0	0	0
1	0	+1,682	0	0	2,828	0	0	0	0
1	0	-1,682	0	0	2,828	0	0	0	0
1	0	0	+1,682	0	0	2,828	0	0	0
1	0	0	-1,682	0	0	2,828	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0

Інформаційна матриця $M = F^T F$ має вигляд

$$M = \begin{bmatrix} N & 0 & (2^{n-p} + 2\alpha^2)l_n & 0 \\ 0 & (2^{n-p} + 2\alpha^2)l_n & 0 & 0 \\ (2^{n-p} + 2\alpha^2)l_n & 0 & B_{nn} & 0 \\ 0 & 0 & 0 & 2^{n-p} E_{\binom{n}{2}} \end{bmatrix}. \quad (52)$$

l_n -- вектор-стовпець із елементами 1.

$$B_{nn} = \begin{bmatrix} 2^{n-p} + 2\alpha^4 & 2^{n-p} & \dots & 2^{n-p} \\ 2^{n-p} & 2^{n-p} + 2\alpha^4 & \dots & 2^{n-p} \\ \dots & \dots & \dots & \dots \\ 2^{n-p} & 2^{n-p} & \dots & 2^{n-p} + 2\alpha^4 \end{bmatrix}. \quad (53)$$

Із (52), (53) випливає, що при використанні ротатбельного плану оцінки коефіцієнтів при лінійних членах та парних взаємодіях є некорельовані із оцінками решти коефіцієнтів, а оцінки коефіцієнтів при квадратичних членах корельовані між собою та оцінкою вільного члену. Цей факт дозволяє при статистичному аналізі результатів розрахунку робити висновки про значимість коефіцієнтів при лінійних членах та взаємодіях незалежно від значень решти коефіцієнтів.

Оцінки коефіцієнтів та їхніх дисперсій:

Позначимо:

$$\begin{aligned} \lambda_1 &= \frac{N2^{n-p}}{(2^{n-p} + 2\alpha^2)^2}; \quad \lambda_2 = \frac{N}{2^{n-p} + 2\alpha^2}; \quad A = \frac{1}{2\lambda_1[(n+2)\lambda_1 - n]}. \\ (0\tilde{y}) &= \sum_{j=1}^N \tilde{y}^j, \quad (i\tilde{y}) = \sum_{j=1}^N x_i^j \tilde{y}^j; \quad (il\tilde{y}) = \sum_{j=1}^N x_i^j x_l^j \tilde{y}^j. \\ \hat{a}_i &= \begin{cases} \frac{A}{N} \left[2\lambda_1(n+2)(0\tilde{y}) - 2\lambda_1 \sum_{j=1}^n (jj\tilde{y}) \right], & i = 0, \\ \frac{\lambda_2}{N} (i\tilde{y}), & i = \overline{1, n}. \end{cases} \end{aligned} \quad (54)$$

$$\begin{aligned} \hat{a}_{n+1} &= \frac{A}{N} \left[2\lambda_2^2[(n+2)\lambda_1 - n](ii\tilde{y}) + \lambda_2(1 - \lambda_1) \sum_{j=1}^n (jj\tilde{y}) - 2\lambda_1\lambda_2(0\tilde{y}) \right], \quad i = \overline{1, n}. \\ \hat{a}_{2n+j} &= \frac{\lambda_2^2}{N\lambda_2} (il\tilde{y}), \quad j = \overline{1, K-2n}. \end{aligned} \quad (55)$$

Вочевидь, для дисперсій маємо:

$$s_i^2 = \begin{cases} 2 \frac{A}{N} \lambda_1^2 (n+2) s^2, & i = 0, \\ \frac{\lambda_2}{N} s^2, & i = \overline{1, n}, \\ \frac{A}{N} [\lambda_1 (n+1) - (n-1) \lambda_1^2] s^2, & i = \overline{n+1, 2n}, \\ \frac{\lambda_2^2}{N \lambda_1} s^2, & i = \overline{2n+1, K}. \end{cases} \quad (56)$$

$s^2 = s^2/\nu$ -- оцінка дисперсії значень $\tilde{y}$ (дисперсія спостережень) із числом ступенів свободи

$$\varphi_2 = (2^{n-p} + 2n)(\nu - 1) + \nu n_0 - 1.$$

Перевірка адекватності моделі для випадку, коли використовується ротатбельний план. Сума квадратів S_D , що характеризує неадекватність моделі:

$$S_D = n_0 \nu (\tilde{y}^0 - \hat{y}^0)^2 + \nu \sum_{j=1}^{N_1} (\tilde{y}^j - \hat{y}^j)^2. \quad (57)$$

Тут $\hat{y}^j$ -- значення вихідної змінної у точці плану x^j , обчислені за допомогою отриманого рівняння; $\tilde{y}^0$ -- середнє значення залежної змінної у центрі плану (50) Із сумою квадратів в (57) зв'язане число ступенів свободи

$$\varphi_1 = N - (K + 1) - (n_0 - 1) = N - K - n_0. \quad (58)$$

Модель вважається адекватною при обраному рівні значимості $\alpha = 1 - p$, якщо

$$F = \frac{S_D/\varphi_1}{S_e/\varphi_2} < F_{cr}. \quad (59)$$

Оцінка значимості коефіцієнтів моделі виконується порівнянням t_{cr} із таблиці для t -розподілів та значенням $|\tilde{y}^0 - \hat{a}^0|$. $|\hat{a}_i| \sim s_i^2 t_{cr}$.

4.8 Побудова канонічної моделі

для квадратичної регресійної моделі

Розглянемо n -факторну модель другого порядку

$$y(x) = \beta_0 + \sum_{i=1}^n (\beta_i x_i + \beta_{ii} x_i^2) + \sum_{i < j} \beta_{ij} x_i x_j, \quad (60)$$

де x_i -- кодовані значення факторів:

$$x_i = \frac{X_i - X_{i0}}{\Delta X_i}, \quad \Delta X_i = (X_{i\max} - X_{i\min})/2, \quad (61)$$

X_i -- значення факторів у натуральних змінних, X_{i0} -- координати центру експерименту

$$X_{i0} = (X_{i\max} + X_{i\min})/2, \quad (62)$$

Передбачається, що:

а) сформульована мета і завдання роботи у техніко-економічних термінах, зібрана і проаналізована апріорна інформація, прийнято рішення про проведення алгоритмізованого експерименту для отримання поліноміальної моделі;

б) обрано r критеріїв якості роботи об'єкту Y_j (параметрів виходу системи або параметрів оптимізації) і n основних факторів X_i , які визначають змінювання виходу Y_j ; визначено положення центру експерименту X_{i0} і піддіапазон варіювання ΔX_i за кожним i -м фактором;

в) на ґрунті критеріїв оптимальності планів вибрана матриця експерименту і схема організації вимірювань;

г) реалізовано експеримент у N точках плану, визначена оцінка похибки експерименту s_0 , за методом найменших квадратів розраховано оцінки коефіцієнтів і отримана модель

$$\hat{Y} = b_0 + \sum_{i=1}^n (b_i x_i + b_{ii} x_i^2) + \sum_{i<j} b_{ij} x_i x_j, \quad (63)$$

д) оскільки оцінки b_0, b_i, b_{ij} , що визначаються за вибіркою із N дослідів, є випадкові величини, розподілені за певним законом, висувається гіпотеза про відмінність цих коефіцієнтів від нуля, про адекватність моделі експериментальним даним, про її інформаційну спроможність тощо, які перевірено за відповідними критеріями Cr при обраному ризику α помилкового відхилення цих гіпотез (рівень значимості, або похибка першого роду).

Тепер потрібно прийняти техніко-економічне рішення.

Один із наочних підходів до прийняття рішень на ґрунті повної поліноміальної моделі є побудова (визначення) геометричної фігури, що описує модель (118). Як відомо із курсу аналітичної геометрії, для того, щоб отримати канонічну форму для квадратичної моделі другого порядку, треба виконати перетворення координат (факторів), тобто паралельний перенос вздовж осі координат і поворот осей координат для того, щоб зникли взаємодії факторів (тобто складова $\sum_{i<j} b_{ij} x_i x_j$).

Будемо розглядати трьохфакторну модель (118) (для простоти). Розв'язання цієї задачі складається із таких кроків:

1) Запис моделі у стандартній формі загального рівняння поверхні другого порядку:

$$a_0 + \sum_{i=1}^n (2a_i x_i + a_{ii} x_i^2) + 2 \sum_{i<j} a_{ij} x_i x_j = 0 \quad (64)$$

або

$$(b_0 - \hat{Y}) + \sum_{i=1}^n (2(b_i/2) x_i + a_{ii} x_i^2) + 2 \sum_{i<j} (b_{ij}/2) x_i x_j = 0 \quad (65)$$

2) Визначення інваріантів I поверхні (120). Позначимо

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}.$$

$$I_1 = b_{11} + b_{22} + b_{33}.$$

$$I_2 = \begin{vmatrix} b_{11} & b_{12}/2 \\ b_{21}/2 & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{13}/2 \\ b_{31}/2 & b_{33} \end{vmatrix} + \begin{vmatrix} b_{22} & b_{23}/2 \\ b_{32}/2 & b_{33} \end{vmatrix}$$

$$= b_{11}b_{22} + b_{22}b_{33} + b_{11}b_{33} - (b_{12}^2 + b_{13}^2 + b_{23}^2)/4.$$

$$I_3 = \begin{vmatrix} b_{11} & b_{12}/2 & b_{13}/2 \\ b_{21}/2 & b_{22} & b_{23}/2 \\ b_{31}/2 & b_{32}/2 & b_{33} \end{vmatrix}$$

$$= b_{11}b_{22}b_{33} + b_{12}b_{13}b_{23}/4 - (b_{13}^2b_{22} + b_{12}^2b_{33} + b_{23}^2b_{11})/4.$$

За знаками інваріантів I_1, I_2, I_3 визначаємо тип поверхні другого порядку. Якщо всі $I_i > 0$, поверхня -- еліпсоїд.

3) Координати центру поверхні визначаємо за формулами (у матриці B послідовно заміщуємо відповідні стовпці і обчислюємо визначники):

$$x_{1s} = \frac{1}{I_3} \begin{vmatrix} -b_1/2 & b_{12}/2 & b_{13}/2 \\ -b_2/2 & b_{22} & b_{23}/2 \\ -b_3/2 & b_{32}/2 & b_{33} \end{vmatrix}.$$

$$x_{2s} = \frac{1}{I_3} \begin{vmatrix} b_{11} & -b_1/2 & b_{13}/2 \\ b_{12}/2 & -b_2/2 & b_{23}/2 \\ b_{13}/2 & -b_3/2 & b_{33} \end{vmatrix}.$$

$$x_{3s} = \frac{1}{I_3} \begin{vmatrix} b_{11} & b_{12}/2 & -b_1/2 \\ b_{12}/2 & b_{22} & -b_2/2 \\ b_{13}/2 & b_{23}/2 & -b_3/2 \end{vmatrix}.$$

4) Визначаємо значення $\hat{Y}_s$ у центрі еліпсоїду підстановкою x_{is} у (63).

5) Визначаємо коефіцієнтів канонічного рівняння поверхні:

$$\hat{Y} - \hat{Y}_s = B_1 \hat{X}_1^2 + B_2 \hat{X}_2^2 + B_3 \hat{X}_3^2. \quad (66)$$

де B_i -- корені характеристичного рівняння:

$$\lambda^3 - I_1 \lambda^2 + I_2 \lambda - I_3 = 0. \quad (67)$$

6) Зв'язок між старими x_i і новими $\hat{X}_i^2$ змінними (координатами) описується системою рівнянь:

$$\left. \begin{aligned} \hat{X}_1 &= l_1(x_1 - x_{1s}) + l_2(x_2 - x_{2s}) + l_3(x_3 - x_{3s}) \\ \hat{X}_2 &= m_1(x_1 - x_{1s}) + m_2(x_2 - x_{2s}) + m_3(x_3 - x_{3s}) \\ \hat{X}_3 &= n_1(x_1 - x_{1s}) + n_2(x_2 - x_{2s}) + n_3(x_3 - x_{3s}) \end{aligned} \right\} \quad (68)$$

Тут $l_i = \cos \alpha_i$, $m_i = \cos \beta_i$, $n_i = \cos \gamma_i$ -- напрямні косинуси кутів, утворених поворотом осі координат, причому

$$l_1^2 + l_2^2 + l_3^2 = 1, \quad m_1^2 + m_2^2 + m_3^2 = 1, \quad n_1^2 + n_2^2 + n_3^2 = 1.$$

Значення напрямних косинусів для кожного кореня λ_i визначаються із системи рівнянь

$$\left. \begin{aligned} (b_{11} - \lambda_i)l_i + b_{12}/2m_i + b_{13}/2n_i &= 0, \\ b_{12}/2l_i + (b_{22} - \lambda_i)m_i + b_{23}/2n_i &= 0, \\ b_{13}/2l_i + b_{23}/2m_i + (b_{33} - \lambda_i)n_i &= 0 \end{aligned} \right\} \quad (69)$$

Після підстановки отриманого розв'язку цієї системи рівнянь відносно напрямних косинусів у (68) отримаємо вирази, що пов'язують нові і старі координати.

Описаний алгоритм надає можливість наочного уявлення про отримані регресійні моделі у вигляді канонічних поверхонь другого порядку.

4.9 Послідовне планування

При використанні планів типу 2^k спочатку слушно ставити невеликі групи дослідів -- мінімально можливі плани типу 2^{k-q} із надією виявити очікувані ефекти. Якщо результати таких дослідів виявляться недостатніми, слід додати наступну групу дослідів (блок 2^{k-q} з метою уточнення коефіцієнтів моделі і визначення її нових складових тощо). У цій ситуації зручно виконувати "поточне" оцінювання коефіцієнтів та їхній аналіз після реалізації кожного додаткового дослідів.

Нехай для визначення поліноміальної моделі із l коефіцієнтами використовується експериментальна програма із m блоків типу 2^{k-q} по $N_w \geq l$, яка дає згідно МНК ортогональні оцінки коефіцієнтів l із мінімальними дисперсіями $s_y^2/(mN_w)$. Тоді для "поточного" оцінювання після включення кожного нового дослідів (або групи дослідів) справджується вираз

$$B^* = B + \frac{R^T(Y - \hat{Y})}{mN_w + l},$$

де B^* -- l -вектор вхідних оцінок b , що отримано на основі m початкових блоків 2^{k-q} ; B -- l -вектор нових оцінок із урахуванням g доданих дослідів із нового блоку; R -- $(g \times l)$ -матриця доданих дослідів r_k ($k = 1, 2, \dots, g = N_w$); Y -- g -вектор нових спостережень; $\hat{Y} = RB$ -- вектор передбачених значень відгуку у нових дослідів.

В алгебраїчній формі маємо:

$$B^* = B + \Delta B + \frac{1}{mN_w + l} \sum_{k=1}^g (y_k - \hat{y}_k) r_k,$$

де ΔB -- корекція для k -го додаткового дослідів.

Це рівняння має назву "прогноз-корекція". Користуватися цим рівнянням можна тільки коли:

- а) усі m блоків разом утворюють план, ортогональний у межах усього експерименту;
- б) усі рядки кожного блоку у межах вибраного розміру N_w ($N_w \geq l$) взаємно ортогональні:

$$r_k \cdot r_\lambda = 0, \quad k \neq \lambda.$$

Ці умови завжди виконуються, якщо блоками є 2^k або 2^{k-q} плани, для яких $X^T X = N_w \cdot I_N$ і $RR^T = N_w I_g$.

Дисперсія кожного коефіцієнту b^* при цьому буде

$$s^2[b^*] = \frac{1}{mN_w} \left[1 - \frac{g}{mN_w + l} \right] s^2[y].$$

Загальна сума квадратів

$$s_y^2 = \sum_{k=1}^{mN_w} y_k^2, \quad f = mN_w.$$

Сума квадратів, що пов'язана із коефіцієнтами регресії:

$$s_y^2 = \sum_{k=1}^{mN_w} \hat{y}_k^2, \quad f = l.$$

Залишкова сума квадратів:

$$s_R^2 = \sum_{k=1}^{mN} (y_k - \hat{y}_k)^2, \quad f = mN_w - l.$$

f -- число ступенів свободи.

У разі коли модель вибраної структури адекватна, оцінкою залишкової дисперсії s_R^2 буде

$$\frac{s_R^2}{mN_w - l}.$$

Додавання кожного нового досліду змінює залишкову суму квадратів s_R^2 на:

$$\Delta s_R^2 = \frac{mN_w}{mN_w + l} (y_k - \hat{y}_k)^2.$$

Якщо число коефіцієнтів $l < N_w$, у модель додають додаткові взаємодії, щоб виконати умову $N_w = l$.

Із використанням послідовного методу можна виконувати оцінку коефіцієнтів моделі та аналіз їх значимості за критерієм Ст'юдента після кожного додаткового дослідів. Коефіцієнт b буде значимим, якщо його величина перевищує довірчий інтервал:

$$\Delta b = t \cdot s[b],$$

де t -- значення критерію Ст'юдента для вибраного рівня значимості α і числа ступенів свободи f , що відповідає числу проведених дослідів; $s[b]$ -- "поточне" значення похибки у визначенні коефіцієнтів регресії.

Зі збільшенням кількості нових точок r відношення $b/s[b]$ збільшується за рахунок зменшення $s[b]$, тобто стається відокремлення сигналу на фоні шуму. Із порівняння "поточного" значення коефіцієнту b із довірчим інтервалом можна зупинити досліді після виконання умови $|b| > \Delta b$.

Зауважимо, що включення додаткових дослідів у такий спосіб не еквівалентно простому зменшенню похибки, а надає нову інформацію, що укрупнює структуру моделі у процесі експериментування. Якщо, наприклад, реалізована репліка 2^{3-1} із $I = x_1 x_2 x_3$, а потім у результаті послідовного додавання дослідів -- уся репліка 2^{3-1} із $I = -x_1 x_2 x_3$, за даними ітогового експерименту 2^3 можна обчислити усі двохфакторні взаємодії і взаємодію третього порядку.

Метод послідовної оцінки параметрів слушно застосовувати у таких випадках:

а) при обробці результатів багатостадійного експерименту великого розміру, коли для зручності аналізу та його оперативності шукані оцінки раціональніше обчислювати послідовно на кожному кроці отримання нових даних;

б) для економного та гнучкого експериментування за різних обмежень. Зокрема, коли вартість дослідів велика, а інтервали часу між дослідями значні, можна отримати необхідну інформацію до закінчення всієї програми дослідів і тим самим скоротити витрати на їх проведення;

в) при аналізі експерименту за неповними даними у випадку втрати результатів окремих дослідів, або за наявності результатів (викидів), що підлягають виключенню.

Послідовна схема аналізу планів 2^k слушна, наприклад, при проведенні клінічних експериментів з випробування нових препаратів (нових режимів лікування), коли важливо найскоріше накопичити інформацію для прийняття рішення про перехід на новий, ефективніший препарат (режим лікування).

При цьому можна скоротити число повторних аналізів у групі, що досліджується, або кількість цієї групи для прийняття обгрунтованого рішення, що є суттєвим як з економічних так й етичних міркувань. Така схема експериментування й аналізу відповідає *послідовній процедурі Вальда*. Відомо, що число дослідів за використання послідовної процедури у середньому у два рази менше за число випробувань при використанні класичних процедур тієї самої потужності.

Послідовно спланований експеримент зупиняється, коли досягнуто задану характеристику точності або закінчуються ресурси, що відведено на експеримент.

4.8.1 Лінійна регресія

Для лінійної за параметрами регресії послідовне планування ґрунтується на такому принципі: на кожному кроці -- найбільше уточнення оцінок параметрів у сенсі обраного критерію. У випадку, що розглядається, послідовне планування не має переваги порівняно із статичним за точністю похибок, що отримуються, але йому властива важлива якість:

при послідовному плануванні можна легко проконтролювати точність оцінок і тому природний критерій зупинки -- досягнення заданої точності.

Розглядається модель у вигляді

$$y(x, a) = \sum_{i=0}^k a_i f_i(x) = a^T f(x). \quad (70)$$

Нехай на основі вибірки обсягу N за методом найменших квадратів отримано оцінки a^N , а також обчислено коваріаційну матрицю оцінок (або пропорційну до неї дисперсійну матрицю плану C_N) і залишкову суму квадратів S_N .

Вирази для розрахунку цих величин відрізняються від наведених вище наявністю індексу N , що визначає число реалізованих дослідів:

$$\begin{aligned} a^N &= (F_N^T F_N)^{-1} F_N^T \tilde{Y}_N = C_N F_N^T \tilde{Y}_N, \quad C_N = (F_N^T F_N)^{-1}; \\ S_N &= (\tilde{Y}_N - F_N a^N)^T (\tilde{Y}_N - F_N a^N) = \tilde{Y}_N^T \tilde{Y}_N - (a^N)^T F_N^T \tilde{Y}_N. \end{aligned} \quad (71)$$

Матриця F_N має вигляд

$$\begin{bmatrix} f_0(x^1) & \dots & f_k(x^1) \\ f_0(x^2) & \dots & f_k(x^2) \\ \dots & & \\ f_0(x^N) & \dots & f_k(x^N) \end{bmatrix} \quad (73)$$

Додамо тепер до вибірки обсягу N ще m експериментів, які визначаються

матрицями

$$X_m = \begin{bmatrix} x_1^{N+1} & x_2^{N+1} & \dots & x_n^{N+1} \\ x_1^{N+2} & x_2^{N+2} & \dots & x_n^{N+2} \\ \dots & \dots & \dots & \dots \\ x_1^{N+m} & x_2^{N+m} & \dots & x_n^{N+m} \end{bmatrix};$$

$$F_m = \begin{bmatrix} f_0(x^{N+1}) & f_1(x^{N+1}) & \dots & f_k(x^{N+1}) \\ f_0(x^{N+2}) & f_1(x^{N+2}) & \dots & f_k(x^{N+2}) \\ \dots & \dots & \dots & \dots \\ f_0(x^{N+m}) & f_1(x^{N+m}) & \dots & f_k(x^{N+m}) \end{bmatrix}; \tilde{Y}_m = \begin{bmatrix} \tilde{y}^{N+1} \\ \tilde{y}^{N+2} \\ \dots \\ \tilde{y}^{N+m} \end{bmatrix}. \quad (75)$$

Вважається, що дисперсія похибок спостережень у нових експериментах та сама, що й у попередніх. Треба побудувати алгоритми обчислень:

а) оцінок a^{N+m} через оцінку a^N і додаткову інформацію, що міститься у матрицях F_m та $\tilde{Y}_m$;

б) дисперсійної матриці C_{N+m} через матрицю C_N та F_m ;

в) залишкової суми квадратів S_{N+m} через S_N , a^N и $\tilde{Y}_m$.

Враховуючи (71), отримуємо:

$$a^{N+m} = C_{N+m}^{-1} F_{N+m}^T \tilde{Y}_{N+m} = (F_N^T F_N + F_m^T F_m)^{-1} (F_N^T \tilde{Y}_N + F_m^T \tilde{Y}_m). \quad (76)$$

Оскільки

$$(F_N^T F_N + F_m^T F_m)^{-1} = C_N - C_N F_m^T [F_m C_N F_m^T + E_m]^{-1} F_m C_N,$$

то

$$C_{N+m} = C_N - C_N F_m^T [F_m C_N F_m^T + E_m]^{-1} F_m C_N. \quad (77)$$

Замість (77) у (76) і виконаємо перетворення, отримаємо рекурентне співвідношення для оцінок коефіцієнтів

$$a^{N+m} = a^N + C_N F_m^T [F_m C_N F_m^T + E_m]^{-1} (\tilde{Y}_m - F_m a^N). \quad (78)$$

Застосування виразів (77), (78) при послідовній побудові математичної моделі призводить до простіших обчислень.

Найпростіший вигляд набувають обчислювальні формули у випадку, коли $m = 1$, тобто перерахунок коефіцієнтів виконується після кожного досліді:

$$C_{N+1} = C_N - \frac{C_N f(x^{N+1}) f^T(x^{N+1}) C_N}{1 + f^T(x^{N+1}) C_N f(x^{N+1})},$$

$$a^{N+1} = a^N + \frac{C_N f(x^{N+1})}{1 + f^T(x^{N+1}) C_N f(x^{N+1})} [\tilde{y}^{N+1} - f^T(x^{N+1}) a^N]. \quad (80)$$

Якщо $N = 0$, тобто треба обчислити оцінки для першої серії дослідів, величини C_0 і a_0 необхідно якимось чином задати. У цьому випадку можна діяти так: реалізувати план, що складається із $N \geq k+1$ експериментів; для цього плану знайти матрицю C_N і оцінки a^N . Після цього стає можливим застосування наведених вище рекурентних формул.

Можливий й інший спосіб. Задамося значеннями

$$a^0 = 0; \quad C_0 = E_{k+1} c^2,$$

де c -- велике число ($c \geq 10^{15}$) і застосуємо рекурентні формули, починаючи з першої серії дослідів. Тоді після i -ї серії із m дослідів, $im \geq k+1$, оцінки коефіцієнтів a_*^{mi} практично не будуть відрізнятися від строго оптимальних оцінок a^{mi} , що обчислені за формулою (129).

Запишемо очевидний вираз для залишкової суми квадратів:

$$S_{N+m} = (\tilde{Y}_N - F_N a^{N+m})^T (\tilde{Y}_N - F_N a^{N+m}) + (\tilde{Y}_m - F_m a^{N+m})^T (\tilde{Y}_m - F_m a^{N+m}). \quad (81)$$

Враховуючи ((190)) і виконуючи перетворення, аналогічні до ((188)), отримаємо

$$S_{N+m} = S_N + (\tilde{Y}_m - F_m a^N)^T (E_m + F_m C_N F_m^T)^{-1} (\tilde{Y}_m - F_m a^N). \quad (82)$$

У разі, коли $m = 1$, тобто додано тільки одну експериментальну точку, маємо

$$S_{N+1} = S_N + \frac{1}{1 + f^T(z^{N+1})C_N f(x^{N+1})} [\tilde{y}^{N+1} - f^T(x^{N+1})a^N]^2. \quad (83)$$

Залишкова сума квадратів може використовуватися для обчислення залишкової дисперсії за формулою

$$(s_R^2)^{N+m} = \frac{S_{N+m}}{N+m-(k+1)}. \quad (84)$$

Залишкова дисперсія є оцінка дисперсії похибок спостережень із числом ступенів свободи $\varphi = N+m-(k+1)$.

У разі коли похибка спостережень e розподілена за нормальним законом, оцінку $(s_R^2)^{N+m}$ можна використовувати при перевірці різного роду гіпотез про коефіцієнти моделі, а також при побудові довірчих інтервалів для величин, що оцінюються.

Приклад.

Розглянемо модель процесу, яку отримано експериментальним шляхом і яка зв'язує вихід процесу із такими змінними:

z_1 -- доза каталізатора; z_2 -- відношення реагентів; z_3 -- час реакції.

Залежність між цими змінними шукалась у вигляді

$$y(a, x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3.$$

$$x_1 = \frac{z_1 - 4}{1}; \quad x_2 = \frac{z_2 - 1,4}{0,1}; \quad x_3 = \frac{z_3 - 3}{0,5}.$$

Спочатку було реалізовано насичений план, що являв собою піврепліку повного факторного експерименту із генератором $x_2 = -x_1 x_3$.

Статистичний аналіз отриманих результатів засвідчив, що коефіцієнт a_2 незначимо відрізняється від нуля. Було зроблено припущення, що причина отримання такого результату криється у занадто великій дисперсії похибок спостережень. Після додавання двох дослідів, що взяті з піврепліки повного факторного експерименту із генератором $x_3 = x_1 x_2$, отримано значимі величини оцінок усіх коефіцієнтів і досліди були зупинені.

Оцінка s^2 дисперсії значень $\tilde{y}$, що визначена за рядом паралельних дослідів при числі ступенів свободи $\varphi_2 = 4$, дорівнює $s^2 = 10,9$. Дисперсійна матриця після перших чотирьох дослідів має вигляд

$$C_4 = \frac{1}{4} E_4.$$

Оцінки коефіцієнтів отримано за формулою

$$\hat{a}_i^4 = \frac{1}{4} \sum_{j=1}^4 \tilde{y}^j x_i^j; \quad (s_i^2)^4 = \frac{s^2}{4} = 2,73.$$

Перевірка значимості коефіцієнтів виконувалася за допомогою критерію Ст'юдента для рівня значимості $\alpha = 0,1$. Після j -го експерименту залишкова дисперсія визначається за формулою

$$(s_R^2)^j = \frac{S_j}{j - (k + 1)} = \frac{S_j}{j - 4}.$$

Дисперсійна матриця після п'ятого дослідів:

$$C_5 = \frac{1}{32} \begin{bmatrix} 7 & 1 & 1 & -1 \\ 1 & 7 & -1 & 1 \\ 1 & -1 & 7 & 1 \\ -1 & 1 & 1 & 7 \end{bmatrix}.$$

Вектор коефіцієнтів a^5 розраховуємо за формулою (78), а залишкову суму квадратів за формулою (83):

$$a^5 = \begin{bmatrix} 67,77 \\ 15,72 \\ 2,97 \\ 4,07 \end{bmatrix} + \frac{1}{8} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} (58,2 - 53,15) = \begin{bmatrix} 68,4 \\ 15,09 \\ 2,34 \\ 4,70 \end{bmatrix}.$$

$$S_5 = \frac{1}{2} (58,2 - 53,15)^2 = 12,75.$$

Дисперсії оцінок коефіцієнтів дорівнюють

$$(s_i^2)^5 = \frac{7}{32} s^2 = 2,38.$$

Оцінки коефіцієнтів після шостого дослідів визначаються за такими обчисленнями:

$$a^6 = \begin{bmatrix} 68,4 \\ 15,09 \\ 2,34 \\ 4,70 \end{bmatrix} + \frac{1}{64} \begin{bmatrix} 7 & 1 & 1 & -1 \\ 1 & 7 & -1 & 1 \\ 1 & -1 & 7 & 1 \\ -1 & 1 & 1 & 7 \end{bmatrix} \cdot (67,9 - 76,45) = \begin{bmatrix} 67,33 \\ 14,02 \\ 3,41 \\ 5,77 \end{bmatrix}.$$

Залишкова сума квадратів:

$$S_6 = 12,75 + \frac{1}{2} (67,9 - 75,45)^2 = 24,68.$$

Дисперсійна матриця

$$C_6 = \frac{1}{32} \begin{bmatrix} 6 & 0 & 2 & 0 \\ 0 & 6 & 0 & 2 \\ 2 & 0 & 6 & 0 \\ 0 & 2 & 0 & 6 \end{bmatrix}.$$

Звідси, дисперсія оцінок коефіцієнтів

$$(s_i^2)^6 = \frac{6}{32} s^2 = 2,04.$$

4.8.2 Метод Йетса

Метод Йетса використовується для розрахунку коефіцієнтів регресії.

Перша операція полягає у попарному додаванні та відніманні значень відгуку (параметри оптимізації), причому верхнє значення віднімається із нижнього.

Друга операція полягає у тій самій дії, але із числами, що отримано на першій операції, тощо.

Якщо тепер числа після останньої операції поділити на число дослідів, отримаємо значення коефіцієнтів. Операції додавання і віднімання повторюються стільки разів, скільки є факторів.

Табл.4.7 Матриця розрахунків і результат

№	x_0	x_1	x_2	x_1x_2	y	II	III
1	+1	-1	-1	+1	95	185	352
2	+1	+1	-1	-1	90	167	-8
3	+1	-1	+1	-1	85	-5	-18
4	+1	+1	+1	+1	82	-3	2
b	88,0	-2,0	-4,5	0,5			

Зі збільшенням числа факторів число можливих взаємодій швидко росте.

Розглянемо повний факторний експеримент 2^3 .

Табл.4.7. Матриця ПФЕ 2^3

№	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	+	-	-	+	-	-	-	+	y_1
2	+	+	-	-	-	-	+	+	y_2
3	+	-	+	-	-	+	-	+	y_3
4	+	+	+	+	+	+	+	+	y_4
5	+	-	-	-	+	+	+	-	y_5
6	+	+	-	+	-	+	-	-	y_6
7	+	-	+	+	-	+	+	-	y_7
8	+	+	+	-	+	-	-	-	y_8

Повна кількість усіх можливих ефектів, разом із b_0 , лінійними ефектами та взаємодій усіх порядків, дорівнює числу дослідів повного факторного експерименту.

$$C_k^m = \frac{k!}{m!(k-m)!},$$

де k -- число факторів, m -- число елементів у взаємодії. Для плану 2^4 число взаємодій дорівнює 6.

$$C_4^2 = \frac{4!}{2!2!} = 6.$$

Ортогональність матриці планування надає можливість отримувати незалежні оцінки коефіцієнтів. Але це справджується лише у тому разі, коли модель містить тільки лінійні ефекти та ефекти взаємодії. Між тим суттєвими можуть виявитися коефіцієнти при квадратах факторів, їх кубах тощо. Наприклад,

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2.$$

Спроба побудови вектор-стовпців для x_1^2 , x_2^2 призводить до отримання одиничних стовпців, що збігаються один із одним та зі стовпцем x_0 . У цьому випадку говорять, що має місце змішана оцінка. Символічно це записується як

$$b_0 \rightarrow \beta_0 + \sum_{j=1}^k \beta_{jj},$$

де b_0 -- обчислений коефіцієнт, а грецькими буквами позначено невідомі істинні значення вільного члену (β_0) та квадратичних коефіцієнтів (β_{jj}).

Відносно квадратичної моделі для двох факторів отримується така система змішування:

$$b_0 \rightarrow \beta_0 + \sum_{j=1}^k \beta_{jj}, \quad b_2 \rightarrow \beta_2, \quad b_1 \rightarrow \beta_1, \quad b_{12} \rightarrow \beta_{12}.$$

Отже, оцінки всіх коефіцієнтів, крім b_0 , не змішані.

Експеримент, в якому реалізуються усі можливі сполучення рівнів, називають повним факторним експериментом. Якщо число рівнів дорівнює двом, це -- повний факторний експеримент 2^k .

Повний факторний експеримент типу 2^k має властивості симетричності, нормування, ортогональності, ротатбельності (для лінійної моделі).

Коефіцієнти, що обчислено за результатами експерименту, вказують на силу впливу факторів. Ефект фактору дорівнює подвоєному коефіцієнту. Коли ефект одного фактора залежить від рівня, на якому міститься інший фактор, говорять про наявність ефекту взаємодії двох факторів.

Із повного факторного експерименту не можна отримати інформацію про квадратичні члени. Оскільки у повному факторному експерименті різниця між числом дослідів і числом коефіцієнтів велика, виникає проблема зменшення числа дослідів.

4.10 Багаторівневі і несиметричні факторні плани

За аналогією з планами 2^k можна розглядати багаторівневі факторні плани типу p^k із числом рівнів $p > 2$, а також плани із різним числом рівнів для кожного фактору. Використання таких планів є досить типовим для медико-біологічних досліджень. Це пов'язано із тим, що поверхні відгуку об'єктів, що досліджуються, мають складні конфігурації; для їхнього опису необхідно використовувати поліноми досить високого порядку q , які потребують великої кількості рівнів для факторів $p \geq q+1$. Характерна особливість майже будь-якого медико-біологічного експерименту є необхідність урахування великої кількості якісних факторів із різним числом рівнів та балансування

різних неоднорідностей.

Наприклад, при оптимізації процесів культивування мікроорганізмів та вірусів поряд із дослідженням кількісних факторів (вік культури, доза зараження тощо) приходиться аналізувати якісні фактори, таких як вид агарового покриття, тип, штаму, що використовується тощо. Такі постановки задач урахування якісних факторів та виключення впливу неоднорідностей трапляються також у клінічних та інших дослідженнях. Вибір кількості і значень рівнів факторів у таких задачах визначається видом моделі та існуючими обмеженнями на кількість дослідів.

Іншим природним критерієм оптимальності при виборі планів зазвичай є критерій ортогональності плану, який забезпечує простоту аналізу та інтерпретації даних.

Розглянемо плани для довільної кількості k факторів (кількісних і якісних) із довільним числом рівнів p_i для кожного з них. Такий план із N дослідів записується як $p_1 \times p_2 \times \dots \times p_k // N$ або просто $p_1 \times p_2 \times \dots \times p_k$.

4.9.1 Загальні поняття про факторні плани та моделі

Щоб визначити систему змішування, запроваджено поняття *співвідношень, що генерують, контрастів та роздільної властивості репліки*.

Співвідношення, що генерують (ГС), показують, які із взаємодій замінені новими факторами. Наприклад, для трьохфакторного експерименту x_1, x_2, x_3 $x_4 = x_1 x_2$ є генеруючим співвідношенням.

Визначальним контрастом (ВК) називають співвідношення, що визначається як добуток стовпців плану, яке дає ± 1 : $I = x_i x_j \dots x_k = \pm 1$.

Наприклад, для двох півреплік 2^{3-1} із $I = x_1 x_2 x_3$ та $I = -x_1 x_2 x_3$, щоб узнати систему змішування ефектів, досить помножити почергово стовпці $x_1 x_2 x_3$ на $I = \pm x_1 x_2 x_3$.

Роздільна властивість репліки є важливе поняття: воно характеризує можливість оцінки моделі відповідної структури. План із роздільною властивістю R позначається як 2_R^{k-q} , де R відповідає числу факторів у ВК, із мінімального числа таких факторів. Найрозповсюдженішими є плани $R = III, IV, V$, які забезпечують оцінку відповідно: усіх головних ефектів факторів (що змішані із парними взаємодіями); головних ефектів усіх факторів окремо від їхніх парних взаємодій (які змішані між собою); головних ефектів та їхніх парних взаємодій, коли решта факторів вважаються рівними нулю.

Вочевидь, дробові репліки із однаковим числом дослідів можуть бути різної роздільної властивості. Наприклад, для плану 2^4 можлива побудова восьми півреплік з різними R із такими ВК:

$$I = \pm x_1 x_2 x_3; I = \pm x_1 x_2 x_4; I = \pm x_2 x_3 x_4; I = \pm x_1 x_2 x_3 x_4.$$

Якщо треба визначити "чисті" лінійні ефекти, можна вибрати піврепліку 2_{III}^{4-1} із $I = \pm x_1 x_2 x_3$ або $I = \pm x_1 x_3 x_4$, тобто із генеруючими співвідношеннями $x_4 = x_1 x_2$ або $x_4 = x_1 x_3$. Число членів у ВК репліки $R = II$.

Якщо потрібно, щоб лінійні ефекти оцінювалися незалежно від парних взаємодій, слід використовувати піврепліки 2^{4-1}_{IV} , які мають ВК $I = x_1 x_2 x_3 x_4$ та $I = -x_1 x_2 x_3 x_4$ із числом членів в ВК $R = IV$, тобто ГС типу $x_4 = \pm x_1 x_2 x_3$.

Такі піврепліки, в яких нові фактори, що запроваджуються, прирівнюються взаємодіям найвищих порядків вхідних факторів, мають максимальну роздільну

спроможність R , і називаються головними піврепліками.

Для п'яти факторів головні піврепліки 2^{5-1}_V задаються ВК: $I = \pm x_1 x_2 x_3 x_4 x_5$, тобто ГС типу $x_5 = x_1 x_2 x_3 x_4$ або $x_5 = -x_1 x_2 x_3 x_4$. У таких піврепліках лінійні ефекти змішані із взаємодіями третього порядку, взаємодії першого порядку -- із взаємодіями другого порядку, і можна оцінити незалежно усі b_i b_{ij} , коли решта ефектів є незначні.

Четверть репліки плану 2^{5-2} можна отримати, якщо взяти $x_4 = \pm x_1 x_2 x_3$, а x_5 прирівняти одному із парних взаємодій, наприклад, $x_5 = x_1 x_2$. Це відповідає вибору ВК $I = x_1 x_2 x_3 x_4$ і $I = x_1 x_2 x_5$.

Для $k = 6$ факторів можна побудувати план 2^{6-3}_{III} , якщо прирівняти x_4 потрійній взаємодії, а x_5, x_6 будь-якому із трьох парних взаємодій. Наприклад:

$$I = x_1 x_2 x_3 = x_1 x_2 x_6 = x_1 x_2 x_3 x_4 = x_3 x_4 x_5 = x_2 x_4 x_6 = x_2 x_3 x_5 x_6 = x_1 x_4 x_5 x_6.$$

Приклад.

Досліджується процес отримання білкового збагачувача (використовується при виготовленні продуктів дієтичного харчування) шляхом сумісного осаджування білків крові тварин і обрату за присутності кальцію хлориду. Треба оцінити ефекти впливу 4-х факторів: кількості кальцію хлориду X_1 (г/л), вагового співвідношення крові і обрату X_2 , тривалості теплової обробки суміші крові та обрату X_3 (хв.), температури обробки X_4 ($^{\circ}\text{C}$) на основний вихідний показник -- ступінь використання білків $y(\%)$.

Експеримент проводився у виробничих умовах м'ясокомбінату, і треба було мінімізувати число можливих дослідів. Кінцева мета досліджень полягала у відшукуванні оптимального режиму процесу та отриманні математичного опису області оптимуму. Для вирішення поставленої задачі із використанням апріорної інформації про процес було реалізовано план другого порядку на ґрунті репліки 2^{4-1} із $x_1 = x_2 x_3$, $x_4 = x_1 x_2 x_3 x_4$, який надавав можливість оцінити незалежно ефект температури b_4 та ефекти взаємодії температури з рештою факторів (b_{14}, b_{24}, b_{34}).

В табл. **Ошибка! Источник ссылки не найден.** наведено частку результатів проведеного експерименту, що відповідає плану 2^{4-1} із $N = 8$ дослідів для оцінки шуканих ефектів.

Вибраний план ДФЕ 2^{4-1} надає можливість отримати систему оцінок:

$$b_0 \rightarrow \beta_0 + \sum_i \beta_{ii}, b_1 \rightarrow \beta_1 + \beta_{23}, b_2 \rightarrow \beta_2 + \beta_{13}, b_3 \rightarrow \beta_3 + \beta_{12}.$$

$$b_4 \rightarrow \beta_4, b_{14} \rightarrow \beta_{14}, b_{24} \rightarrow \beta_{24}, b_{34} \rightarrow \beta_{34},$$

в яких ефекти, що цікавлять, та взаємодії оцінюються окремо.

Табл.4.8. Матриця плану ДФЕ 2^{4-1}

Фактори		X_1	X_2	X_3	X_4				y
Рівні:									
0		1,0	7,5	5	95				
(+)		1,5	10	7	100				
(-)		0,5	2,5	3	90				
x_i	x_0	x_1	x_2	x_3	x_4	$x_1 x_4$	$x_2 x_4$	$x_3 x_4$	

Дослід:									
1	+	+	+	+	+	+	+	+	96,6
2	+	+	-	-	+	+	-	-	95,6
3	+	-	+	-	+	-	+	+	94,4
4	+	-	-	+	+	-	-	+	83,0
5	+	+	+	+	-	-	-	-	96,0
6	+	+	-	-	-	-	+	+	95,5
7	+	+	-	-	-	+	-	+	94,0
8	+	-	-	+	-	+	+	-	94,0
<i>b</i>	93,12	1,61	0,46	0,58	-1,23	1,41	1,48	-1,35	

Похибка (для підрахунку використовувалися результати 4 дослідів у центрі плану) $s\{\bar{y}\} = 0,7$, $s\{b_i\} = 0,5$, $f_0 = 4 - 1 = 3$, $\Delta b = t_{0,1} \cdot s\{b\} = t_{0,1}(8) = 2,35 \cdot 0,15 = 1,17$.

Якщо відняти із значення b_0 значення y_0 : $b_0 - y_0 = \sum_i \beta_{ii}$, можна отримати оцінку квадратичних ефектів $\sum_i \beta_{ii}$. Значимість величини $\sum_i \beta_{ii}$ вказує на необхідність включення у рівняння регресії квадратичних ефектів b_{ii} , тобто необхідність добудови експерименту до повного плану другого порядку. Значення різниці $b_0 - y_0$ часто використовують для перевірки адекватності моделі, що отримано за результатами двохрівневого ПФЕ або ДФЕ, і прийняття рішень про необхідність переходу до плану другого порядку.

У даному прикладі велике значення $b_0 - y_0 = 4,08 > \Delta b = 1,17$ саме вказує на необхідність урахування квадратичних ефектів і добудови експерименту до плану другого порядку.

4.9.2 Багаторівневі факторні плани

За аналогією з планами 2^k можна розглядати багаторівневі факторні плани типу p^k із різним числом рівнів для кожного фактору. Використання таких планів є типовим для медико-біологічних досліджень. Це пов'язано із тим, що поверхні відгуку таких об'єктів мають складні конфігурації; для їхнього опису треба використовувати поліноми досить високого порядку q , які потребують великої кількості рівнів для факторів $p \geq q + 1$. Характерна особливість майже будь-якого медико-біологічного експерименту є також необхідність урахування великої кількості якісних факторів із різним числом рівнів та балансування різних неоднорідностей.

Наприклад, при оптимізації процесів культивування мікроорганізмів та вірусів поряд із дослідженням кількісних факторів (вік культури, доза зараження тощо) необхідно оцінювати вплив якісних факторів таких, як вигляд агарового покриття, тип штаму, що використовується, тощо. Подібні постановки задач урахування якісних факторів та виключення впливу неоднорідностей наявні також у клінічних та інших дослідженнях.

Іншим природним критерієм оптимальності при виборі планів зазвичай є критерій ортогональності плану, що забезпечує простоту аналізу та інтерпретації даних.

План, що містить всілякі сполучення рівнів усіх факторів із $N = \prod_{j=1}^k p_j$, називається планом повного факторного експерименту, а план, що містить тільки частку цих

сполучень -- дробовим планом.

Позначимо p_j рівнів змінної X_k , що досліджується (якісної або кількісної) через $0, 1, 2, \dots, (p_j - 1)$, а матрицю плану запишемо як:

$$\pi = \begin{bmatrix} x_{11} & \dots & x_{k1} \\ \vdots & & \vdots \\ x_{1N} & \dots & x_{kN} \end{bmatrix},$$

де x_{ij} -- значення i -го фактора у j -му досліді.

План називається **симетричним**, якщо кожний його фактор має однакову кількість рівнів $p_j = p$. Вочевидь, плани 2^k є симетричні, а їхні матриці у прийнятих позначеннях міститимуть елементи x_{ij} , що дорівнюють 0 або 1.

У **несиметричному** плані фактори мають різне число рівнів. Наприклад, план $3 \times 4^2 \times 2^3$ означає, що перший фактор варіюється на трьох рівнях, другий та третій -- на чотирьох рівнях, а решта три фактори -- на двох рівнях кожний.

План називається **рівномірним**, якщо рівні будь-якого фактору з'являються у ньому однакове для даного фактору число разів.

Позначимо n_i^* число появ s -го рівню i -го фактору через n_{ij}^{sg} -- число сумісних появ у рядках плану s -го рівня i -го фактору і g -го рівня j -го фактору ($s = \overline{0, p_i - 1}$; $g = \overline{0, p_j - 1}$; $i, j = \overline{1, k}$; $i \neq j$).

Аналогічно для t факторів частоту сумісних появ їхніх рівнів позначимо через $n_{ij}^{sg\dots r} \dots t$. Факторний план називається **регулярним планом потужності t** , якщо для будь-яких t факторів виконується умова пропорційності частот рівнів:

$$n_{ij}^{sg} = \frac{1}{N} (n_i^s \cdot n_j^g); \quad n_{ijl}^{sgq} = \frac{1}{N^2} (n_i^s \cdot n_j^g \cdot n_l^q); \quad n_{ijlm}^{sgqr} = \frac{1}{N^3} (n_i^s \cdot n_j^g \cdot n_l^q \cdot n_m^r).$$

Число сумісних появ рівнів будь-яких t із k факторів -- комбінацій $x_1, \dots, x_t$, позначатимемо також як $n(x_1, \dots, x_t) = \lambda(x_1, \dots, x_t)$, де $x_i = \overline{0, 1, \dots, p_i - 1}$, тобто $x_i \in P_i$.

Регулярний план потужності t надає можливість отримати модель, що містить ефекти взаємодії факторів (або рівнів факторів у випадку якісних факторів) до порядку $[t/2 - 1]$. Якщо будь-який ефект взаємодії порядку $m < [t/2 - 1]$ відсутній, така модель уже не є моделлю регулярного плану.

Регулярний план потужності t для непарних $t = 2m + 1$ надає можливість незалежно оцінювати усі головні ефекти та взаємодії до m -факторних включно, причому усі вони ортогональні $(m + 1)$ -факторним ефектам взаємодії.

Для кількісних факторів використовують поліноміальну модель, що містить головні ефекти факторів (лінійний, квадратичний тощо) та їх взаємодії.

Головним ефектом фактору X_i називають такий контраст, коефіцієнти якого для усіх спостережень, в яких цей фактор набуває однакових значень, є рівні. Максимальне число незалежних головних ефектів i -го фактору дорівнює $p_i - 1$. Ефектом парної взаємодії i -го і j -го факторів називається такий ортогональний до усіх головних ефектів контраст, коефіцієнти якого для усіх спостережень, в яких ці фактори мають однакові сполучення рівнів, є рівні. Максимальне число ефектів парних взаємодій i -го і j -го факторів дорівнює $(p_i - 1)(p_j - 1)$. Аналогічно визначаються ефекти взаємодій вищих

порядків: $(p_i - 1)(p_j - 1)(p_k - 1)$ тощо.

Для регулярного плану потужності t вектори коефіцієнтів контрастів утворюють матрицю

$$\Theta = [I, \Theta_1, \dots, \Theta_k, \Theta_{12}, \dots, \Theta_{1..k}].$$

де Θ_i -- матриця коефіцієнтів контрастів головних ефектів i -го фактору; θ_{ij} -- матриця коефіцієнтів для ефектів взаємодії i -го і j -го факторів, стовпці якої утворено попарним добутком стовпців матриць Θ_i і Θ_j по одному для кожного фактора тощо. У разі ПФЕ матриця Θ містить у собі стовпець ефектів взаємодії вищого порядку $\Theta_{1..k}$. Усі стовпці такої матриці попарно ортогональні. Зокрема, для планів 2^k матриця коефіцієнтів θ складалася із $N - 1$ попарно ортогональних стовпців, елементами якої були тільки -1 та +1, стовпця $x_0 = +1$ і усі параметри регресійної моделі оцінювалися незалежно.

Для загального випадку факторного плану із $p \geq 2$ рівнями результати експерименту можна подати моделями таких типів:

I. Поліноміальна модель вигляду

$$y = b_0 + \sum_{q=1}^{p_i-1} b_1^{(q)} \Theta_1^{(q)}(x_1) + \dots + \sum_{q=1}^{p_k-1} b_k^{(q)} \Theta_k^{(q)}(x_k) + \Pi + \varepsilon,$$

де $\Theta_i^{(q)}(x_i)$ -- ортогональні поліноми q -го степеню для i -ї змінної, за допомогою яких оцінюються головні ефекти факторів (лінійні, квадратичні, кубічні та інші ефекти). Π відповідає взаємодіям головних ефектів. Зазвичай для кожного фактору функції $\Theta_i^{(q)}(x_i)$ можна задавати системою ортогональних поліномів Чебишова, що визначені у точках $x_{i1}, \dots, x_{iN}$, і тоді указана модель називається чебишовською. У разі повного факторного експерименту член Π містить усі можливі добутки функцій $\Theta_i^{(q)}(x_i)$. У разі повного факторного експерименту член Π містить усі можливі добутки функцій $\Theta_i^{(q)}(x_i)$. Для планів із $R = V$ він містить парні добутки функцій $\Theta_i^{(q)}(x_i)$ та $\Theta_j^{(q)}(x_j)$, по одній для кожного фактору i і j .

Матриці коефіцієнтів регулярного плану оцінюються за МНК незалежно із мінімальними дисперсіями (різними для різних ефектів):

$$b_i^{(q)} = \frac{\sum_{k=1}^N y_k \Theta_i^{(q)}(x_{ik})}{\sum_{k=1}^N [\Theta_i^{(q)}(x_{ik})]^2} = \frac{\sum_{k=1}^N y_k \Theta_{ik}^{(q)}}{\sum_{k=1}^N [\Theta_{ik}^{(q)}]^2};$$

$$s^2[b_i^{(q)}] = \frac{s_0^2}{\sum_{k=1}^N [\Theta_{ik}^{(q)}]^2}; \quad i = \overline{1, k}, \quad q = \overline{1, p_i - 1};$$

$$b_{ij}^{(qs)} = \frac{\sum_{k=1}^N y_k \Theta_{ik}^{(q)} \Theta_{jk}^{(s)}}{\sum_{k=1}^N [\Theta_{ik}^{(q)} \Theta_{jk}^{(s)}]^2}; \quad i \neq j; \quad j = \overline{1, k}; \quad q = \overline{1, p_i - 1}, \quad s = \overline{1, p_j - 1}.$$

$$s^2[b_{i,j}^{qs}] = \frac{s_0^2}{\sum_{k=1}^N [\Theta_{ik}^{(q)} \Theta_{jk}^{(s)}]^2}.$$

Довірчі інтервали для оцінок параметрів визначають звичайним чином:

$$\Delta b_i^{(q)} = f \cdot s[b_i^{(q)}],$$

де f -- значення критерію Ст'юдента для $f = f_0$.

II. Модель ефектів рівнів та взаємодій рівнів факторів:

$$y = \tilde{b}_0 + \sum_{s=0}^{p_i-1} \tilde{b}_i^{(s)} \tilde{x}_1^{(s)} + \dots + \sum_{g=0}^{p_k-1} \tilde{b}_k^{(g)} \tilde{x}_k^{(g)} + \tilde{\Pi} + \varepsilon,$$

де $\tilde{x}_{i(k)}^{(s)} = 1$, якщо i -й фактор в k -му досліді з'являвся на s -му рівні, 0 -- у протилежному випадку; $\tilde{b}_i^{(s)}$ -- ефект s -го рівня i -о фактору; $\tilde{\Pi}$ -- містить члени типу $\tilde{b}_{ij}^{(sg)} \tilde{x}_i^{(s)} \tilde{x}_j^{(g)}$, ..., $\tilde{b}_{ij...k}^{(sg...r)} \tilde{x}_i^{(s)} \tilde{x}_j^{(g)} \dots \tilde{x}_k^{(r)}$; $\tilde{b}_{ij}^{(sg)}$ -- ефект сполучення рівнів s і g факторів X_i і X_j тощо.

Значення $\tilde{x}_i^{(s)}$, $\tilde{x}_i^{(g)}$ тощо у моделі II також відповідає визначенню коефіцієнтів контрастів для головних ефектів і взаємодій-факторів. Але ранг матриці $\tilde{X}$ незалежних змінних менше за число параметрів, що оцінюються. Тому для отримання єдиних МНК-оцінок параметрів b на їхні значення накладають обмеження вигляду

$$\sum_{s=0}^{p_i-1} \tilde{b}_i^{(s)} = 0; \sum_{s=0}^{p_i-1} \tilde{b}_{ij}^{(sg)} = 0; \dots; \sum_{r=0}^{p_i-1} \tilde{b}_{12...k}^{(sg...r)} = 0.$$

У цих позначеннях:

$$y = \tilde{b}_0 + \tilde{b}_1^{(0)} \tilde{x}_1^{(0)} + \dots + \tilde{b}_1^{(p_1-1)} \tilde{x}_1^{(p_1-1)}$$

$$+ \tilde{b}_2^{(0)} \tilde{x}_2^{(0)} + \dots + \tilde{b}_2^{(p_2-1)} \tilde{x}_2^{(p_2-1)} + \tilde{b}_{12}^{(00)} \tilde{x}_1^{(0)} \tilde{x}_2^{(0)} + \dots + \tilde{b}_{12}^{(p_1-1, p_2-1)} \tilde{x}_1^{(p_1-1)} \tilde{x}_2^{(p_2-1)} + \varepsilon$$

за відповідних обмежень на параметри $\tilde{b}$.

За таких обмежень із умови

$$\sum_{k=1}^N (y_k - \hat{y}_k)^2 = \min$$

отримаємо такі вирази для оцінки ефектів рівнів і ефектів взаємодій рівнів у плані із n повторними спостереженнями:

$$b_0 = \frac{1}{N} \sum_{k=1}^N y_k = \frac{G}{p_1 p_2 n}, \quad \tilde{b}_1^{(s)} = \frac{T_{s.}}{p_2 n} - \tilde{b}_0, \quad s = \overline{0, p_1 - 1},$$

$$\tilde{b}_2^{(g)} = \frac{T_{.g}}{p_1 n} - \tilde{b}_0, \quad g = \overline{0, p_2 - 1},$$

$$\tilde{b}_{12}^{(sg)} = \frac{T_{s.g}}{n} - \frac{T_{s.}}{p_2 n} - \frac{T_{.g}}{p_1 n} + \tilde{b}_0, \quad s = \overline{0, p_1 - 1}, \quad g = \overline{0, p_2 - 1}.$$

Тут $T_{s.}, T_{.g}, T_{sg}$ -- суми відгуків по s -му, g -му рівням факторів і комбінаціям цих рівнів. Значення довірчих інтервалів для таких оцінок $\Delta \tilde{b} = ts[\tilde{b}]$, де

$$s\{\tilde{b}_i^{(s)}\} = s_0 \sqrt{\frac{p_i}{n_i^s \cdot n}}; \quad s\{\tilde{b}_{ij}^{(sg)}\} = \frac{s_0}{\sqrt{n^s g_{ij} \cdot n}},$$

де s_0 -- похибка відновлюваності; t -- значення критерію Ст'юдента. За допомогою цих оцінок можна обчислити суми квадратів для факторів та їхніх взаємодій у схемі дисперсійного аналізу.

Факторні моделі I і II еквівалентні, оцінки b і $\tilde{b}$ зв'язані співвідношеннями:

$$\tilde{b}_i^{(s)} = \sum_{q=1}^{p_i-1} b_i^{(q)} \Theta_i^q(s); \quad b_i^{(q)} = \sum_{s=1}^{p_i-1} \tilde{b}_i^{(s)} \Theta_i^{(q)}(s).$$

Тому при аналізі моделі для планів із кількісними і якісними факторами можна обчислювати для усіх факторів як оцінки параметрів b , так і параметрів $\tilde{b}$.

Заключний етап аналізу зводиться до перевірки адекватності отриманих рівнянь. Для цього обчислюють залишкову суму квадратів:

$$S_R = \sum_{k=1}^N (y_k - \hat{y}_k)^2, \quad f_R = N - l - 1,$$

де l -- число значимих ефектів,

$$F = \frac{S_R/f_R}{S_0/f_0}.$$

Якщо $F < F_\alpha$, гіпотеза адекватності рівняння не відкидається.

4.9.3 Вибір оптимальних планів у практичних застосуваннях

Вибір багатофакторного плану (кількість дослідів, число і значення рівнів факторів) треба здійснювати, виходячи з вигляду моделі, що передбачається побудувати, існуючих обмежень на кількість дослідів, вимог оптимальності. Розрахунок потужності експерименту, що планується, до його постановки надає можливість запобігання характерних для прикладних робіт помилок, коли число рівнів, а отже, дослідів або надлишкове або недостатнє для того, щоб зробити висновки про ефекти із заданою надійністю.

Для вирішення розрахунку обсягу вибірки (число рівнів за факторами та кількості повторних спостережень на цих рівнях) за заданою похибкою s_0^2 із числом ступенів свободи $f_0 = 5$ для відокремлення мінімальної значимої різниці γ між значеннями відгуку із потрібною надійністю (похибками $\alpha = 0,05$ та $\beta = 0,90$) можна скористатися такими наближеними формулами:

-- для планів головних ефектів:

$$N \geq (4\rho s/\gamma)^2;$$

-- для планів з m -факторними взаємодіями із числом комірок у таблиці взаємодій g і числом ступенів свободи для чисельника F -статистики, що дорівнює f :

$$N \geq 9s^2(f+1)g/(\gamma^r 2^{m-2});$$

-- для планів поверхні відгуку із повною або неповною моделлю із l і l' членами:

$$N > 36s^2 l(2l-1)/\gamma^2;$$

$$N = 36s^2(l'+1)2l'/\gamma^2.$$

Розглянемо приклади застосування та аналізу ортогональних і квазіортогональних МФП головних ефектів.

Приклад. У табл. 4.10 наведено результати експерименту, що реалізований по

НФП головних ефектів $3^2 \times 2^2 // 9$, який отриманий з СФП $3^4 // 9$ перетворенням двох останніх стовпців. Скорочення: Ф. -- фактори, Р. -- рівні, Д. -- дослід. Вважалось, що залежність y від кількісного фактора X_1 не відрізняється від лінійної, залежність від кількісного фактора X_4 не відрізняється суттєво від квадратичної і взаємодіями між усіма факторами можна знехтувати. Для подання результатів експерименту використовувалася модель типу I вигляду

$$y = b_0 + b_1^{(1)}\theta_1^{(1)} + b_2^{(1)}\theta_2^{(1)} + b_3^{(1)}\theta_3^{(1)} + b_3^{(2)}\theta_3^{(2)} + b_4^{(1)}\theta_4^{(1)} + b_4^{(2)}\theta_4^{(2)} + \varepsilon.$$

$$s^2[b] = \frac{s_0^2}{2} \sum_{k=1}^2 \theta_k^2; \quad s_0^2 = 8,9; \quad f_0 = 18.$$

Кінцеве рівняння, що містить тільки значимі коефіцієнти, має вигляд:

$$\hat{y} = 100,1 + 19,30\theta_2^{(1)} + 8,51\theta_3^{(1)} - 18,10\theta_4^{(2)}.$$

Рівняння адекватно описує дані експерименту, оскільки

$$F = \frac{M[S_R]}{M[S_0]} = \frac{2,18}{8,9} = 0,33 < F_{0,05}(4;18) = 4,1.$$

Із рівняння випливає, що значення виходу y зростатиме при переході фактора X_2 з рівня А на рівень Б (середнє спостереження для рівня Б вище за середнє спостереження для рівня А). Оптимальне значення кількісного фактора X_4 відповідає середньому рівню $x_4 = 1$, тобто $X_4 = 6$, за якого $\theta_4^{(2)} = -2$.

Табл 4.9. План $2^2 \times 3^2$ та ортогональні контрасти для оцінки головних ефектів

Ф.:	X_1	X_2	X_3	X_4	Ортогональні контрасти						y_k	$\hat{y}_k$
Р.:												
0	5	А	І	3								
1	9	В	ІІ	6								
2			ІІІ	9								
x_i	x_1	x_2	x_3	x_4	θ_1	θ_2	θ_3	θ_2^2	θ_4	θ_4^2		
Д.:												
1	0	0	0	0	-1	-1	-1	1	-1	1	54,7	54,2
2	0	1	1	2	-1	2	0	-2	-1	1	120,3	120,6
3	0	0	2	1	-1	-1	1	1	0	2	125,7	125,5
4	1	0	1	1	2	-1	0	-2	0	-2	117,2	117,0
5	1	1	2	0	2	2	1	1	-1	1	129,2	129,1
6	1	0	0	2	2	-1	-1	1	1	1	54,7	54,2
7	0	0	2	2	-1	-1	1	1	1	1	70,7	71,2
8	0	1	0	1	-1	2	-1	1	0	-2	166,2	166,4
9	0	0	1	0	-1	-1	0	-2	-1	1	62,8	62,7
				λ	3	3	1	3	1	3		
				b	0,094	19,3	8,51	0,039	-0,15	-18,1		
				Δb	1,41	1,41	2,44	1,41	2,44	1,41		

4.11 Планування екстремального експерименту

Вирішення різноманітних задач медико-біологічних досліджень пов'язане із оптимізацією, тобто відшукуванням таких значень факторів $x_i = x_i^*$, $i = \overline{1, k}$, за яких забезпечується екстремальне (мінімальне чи максимальне) значення вихідного показника $y^* = f(x_1^*, \dots, x_k^*)$.

Методи, що застосовуються для пошуку екстремуму, називають методами *екстремального експерименту*. Вони передбачають дослідження ділянок поверхні відгуку $y = f(x_1, \dots, x_k)$ шляхом постановки спланованих серій дослідів, кожна із яких надає можливість отримати інформацію про напрямок руху до екстремуму. Отже, екстремальне значення відгуку досягається за допомогою многократної послідовної процедури вивчення поверхні відгуку і просування у факторному просторі.

Існують різні методи оптимізації (градієнтні, методи випадкового пошуку тощо), що відрізняються способами визначення руху та організації саме руху до екстремуму.

У звичайній практиці експериментатор часто використовує метод почергової оптимізації за факторами, відомий як метод Гауса--Зейделя. У такій "класичній" схемі оптимізації послідовно відшукують часткові екстремуми функції відгуку за кожним фактором, за умови фіксованих решти факторів. Очевидна громіздкість цього методу, особливо за великої кількості факторів. За складних форм поверхні відгуку такий метод може призвести до помилкових рішень.

Екстремальним експериментом називають експеримент, спрямований на відшукування екстремуму функції регресії.

Під час вибору методу екстремального планування вирішальним фактором є коштовність експериментів. Часто в якості планів проведення експерименту використовують послідовні плани. За певних припущень ці плани мають асимптотичні властивості збіжності послідовності точок, що отримані, до точки локального екстремуму функції регресії.

4.10.1 Збіжність псевдоградієнтних алгоритмів

Постановка задачі

Нехай $X = R^n$, η -- вимірювана функція на X . Припустимо, що у будь-якій точці $x \in X$ можна обчислювати (вимірювати) реалізацію випадкової величини $y(x) = \eta(x) + \varepsilon(x)$, де випадкові похибки $\varepsilon(x)$ є центровані ($M[\varepsilon(x)] = 0$) і взаємно незалежні за будь-яких $x_1, x_2, \dots$ із X . $\eta(x)$ -- функція регресії, що часто зветься *функцією відгуку*.

Задача планування екстремального експерименту полягає у побудові послідовності точок $\{x_k\}_{k=0}^{\infty} \subset X$, що збігається до точки

$$x^* = \operatorname{argmax}_{x \in X} \eta(x)$$

або такої, що

$$\eta(x_k) = \sup_{x \in X} \eta(x), \quad k \rightarrow \infty.$$

Задача планування екстремального експерименту зазвичай формулюється як задача максимізації. Якщо ставиться задача мінімізації $\tilde{\eta}(x)$ за результатами обчислення

$\tilde{y}(x) = \tilde{\eta}(x) + \tilde{\varepsilon}(x)$, то поклавши

$$\eta(x) = -\tilde{\eta}(x), \quad y(x) = -\tilde{y}(x),$$

приходимо до задачі максимізації.

Іноді при вирішенні задач екстремального планування вдається здійснити вимірювання (із випадковими похибками) градієнту $\nabla \eta$ у точці x :

$$z(x) = \nabla \eta(x) + \zeta(x), \quad \nabla \eta(x) = \left(\frac{\partial \eta(x)}{\partial x(1)}, \dots, \frac{\partial \eta(x)}{\partial x(n)} \right)^T,$$

де $z(x)$, $\zeta(x)$ -- реалізації випадкових векторів розмірності n , $M[\zeta(x)] = 0$, за будь-яких $x_1, x_2, \dots \in X$ випадкові вектори $\zeta_1(x), \zeta_2(x), \dots$ є взаємно незалежні.

Якщо при побудові $\{x_k\}_{k=0}^\infty \subset X$ використовується бодай одна реалізація випадкового вектору z або вимірювання (із випадковою похибкою) старших похідних функції регресії, алгоритм екстремального планування називають *регулярним*. Якщо використовуються тільки значення випадкової функції y , алгоритм називають *пошуковим*.

Розглянемо послідовні алгоритми екстремального планування, які можна подати у такому вигляді:

$$x_{k+1} = x_k + \gamma_{k+1} s_k, \quad (85)$$

де $k \geq 0$ -- номер ітерації, $x_k \in X$, $\gamma_k \geq 0$ -- детермінований множник, $s_k \in X$ -- реалізація випадкового вектору, що залежить від

$$k, x_0, x_1, \dots, x_k, s_0, s_1, \dots, s_{k-1} \quad (86)$$

і реалізацій y або z у наборі точок із X , що визначені спеціальним чином. Вектор s_k часто називають *напрямком руху*, а γ_{k+1} -- *довжиною кроку*. Оскільки $\|s_k\|$ у більшості методів відрізняється від одиниці, справжня довжина кроку у напрямку s_k дорівнюватиме $\gamma_{k+1} \|s_k\|$.

Формула (85) є загальна форма запису послідовних алгоритмів екстремального планування у випадку, коли експерименти спрямовано на відшукування локального екстремуму функції η .

Приклад 1. Нехай вхідний алгоритм має вигляд

$$x_{k+1} = x_k + \Gamma_k z(x_k), \quad (87)$$

де $\Gamma_k : X \rightarrow X$ -- оператор, що залежить від k і $x_0, \dots, x_k$, $z(x_0), \dots, z(x_k)$.

Для того, щоб привести (139) до вигляду (137), досить покласти

$$s_k = \gamma_{k+1}^{-1} \Gamma_k z(x_k),$$

де $\gamma_1, \gamma_2, \dots$ -- послідовність додатних чисел, яку обрано довільним чином.

4.10.2 Планування екстремального експерименту за наявності обмежень

Задача умовної оптимізації. Якщо $X \subset R^n$ -- підмножина, що задана за допомогою деяких обмежень, задачу оптимізації на X називають *задачею умовної оптимізації*. Вибір тієї чи іншої модифікації алгоритму (85) залежить від способу завдання множини X , алгоритму, що використовується, апріорної інформації про функцію η .

Задачі умовної оптимізації зручно розглядати як задачі мінімізації.

Методи штрафних функцій

Методи штрафних функцій є одні із найпростіших і розповсюджених методів вирішення екстремальних задач із обмеженнями. Сутність методу зовнішніх штрафних функцій полягає у приведенні вхідної задачі пошуку інфімуму функції $V(x)$ на множині X до послідовності задач пошуку інфімуму деяких функцій $V_j(x)$ ($j = 1, 2, \dots$) на множині $X_0 \supset X$, а допоміжні функції $V_j(x)$ підбираються так, щоб зі зростанням номера j вони мало відрізнялися від вхідної функції $V(x)$ на множині X і швидко зростали на множині $X_0 \setminus X$.

Нехай зафіксовані деяка множина $X_0 \supset X$ і послідовність функцій $\{P_j(x)\}_{j=1}^{\infty}$, що задані на X_0 . За припущення, що $V(x)$ задана на X_0 , покладемо

$$V_j(x) = V(x) + P_j(x). \quad (88)$$

Послідовність функцій $\{P_j(x)\}_{j=1}^{\infty}$, що визначені і невід'ємні на множині $X_0 \supset X$, називається *послідовністю зовнішніх штрафних функцій* множини X на множині X_0 , якщо

$$\lim_{j \rightarrow \infty} P_j(x) = \begin{cases} 0, & x \in X, \\ +\infty, & x \in X_0 \setminus X. \end{cases} \quad (89)$$

Методом штрафних функцій називається алгоритм вигляду (85), $\gamma_{k+1} \geq 0$, $x_k \in X_0$, s_k -- напрямок на локальний (глобальний) мінімум функції $V_k(x)$ або напрямок псевдоградієнту для функції $V_k(x)$, що визначена за (88).

При плануванні екстремального експерименту досить на k -й ітерації робити лише один крок у напрямку псевдоградієнту функції $V_k(x)$.

Для будь-якої множини $X \subset R^n$ існує і може бути побудована нескінченно велика кількість штрафних функцій. Наприклад, якщо X замкнута, можна покласти

$$P_j(x) = A_j \rho(x), \quad x \in R^n = X_0, \quad j = 1, 2, \dots,$$

де $\{A_j\}$ -- деяка псевдопослідовність додатних чисел, $A_j \rightarrow \infty$ ($j \rightarrow \infty$).

Як правило, при вирішенні конкретних задач оптимізації множина X має вигляд

$$X = \{x \in R^n \mid x \in X_0; \quad g_i(x) \leq 0, \quad i = \overline{1, l}; \quad g_i(x) = 0, \quad i = \overline{l+1, m}\}. \quad (90)$$

де $X_0 \subset R^n$ -- множина простої структури (можливо, $X = R^n$), $X \subset X_0$, Функції $g_i(x)$ ($i = \overline{1, m}$) визначені на X_0 . В якості послідовності штрафних функцій для визначеної таким чином множини X можна обирати

$$P_j(x) = A_j P(x), \quad A_j > 0, \quad j = 1, 2, \dots, \quad A_j \rightarrow \infty, \quad j \rightarrow \infty,$$

$$P(x) = \sum_{i=1}^l [\max\{0, g_i(x)\}]^r + \sum_{i=l+1}^m |g_i(x)|^r, \quad (91)$$

де $x \in X_0$, $r \geq 1$ -- фіксоване число.

На відміну від зовнішніх штрафних функцій внутрішні штрафні функції визначені тільки на множині X .

Нехай множина X замкнута, має внутрішні точки та існує така точка $x \in X$, що множина

$$\{x \in X \mid V(x) \leq V(z)\}$$

компактна. Послідовність неперервних функцій

$$q_i : \text{Int}X \rightarrow R, \quad i = 1, 2, \dots$$

називається *послідовністю внутрішніх штрафних функцій* для X , якщо виконуються умови:

а) $0 < q_{i+1}(x) < q_i(x)$ для всіх $x \in \text{Int}X$ ($i = 1, 2, \dots$);

б) $q_i(x) \rightarrow 0$ при $i \rightarrow \infty$ для всіх $x \in \text{Int}X$;

в) $q_i(x_j) \rightarrow \infty$ при $j \rightarrow \infty$ для будь-якої такої послідовності $\{x_j\} \subset \text{Int}X$, що

$$\lim_{j \rightarrow \infty} x_j \in X \setminus \text{Int}X.$$

При використанні зовнішніх штрафних функцій "штраф платиться" при порушенні умови $x \in X$, а при використанні внутрішніх штрафних функцій -- при наближенні до межі множини X .

Метод внутрішніх штрафних функцій визначається за послідовністю $\{q_i(x)\}_{i=1}^{\infty}$ так саме як метод зовнішніх штрафних функцій за послідовністю $\{P_i(x)\}_{i=1}^{\infty}$.

Метод множників Лагранжа

Розглянемо задачу пошуку мінімуму функції $V(x)$ на множині (90), де функції V, g_i , ($i = \overline{1, m}$) задано на X_0 . Функцією Лагранжа для задачі, що розглядається, називається

$$L(x, \lambda) = V(x) + \sum_{i=1}^m \lambda_i g_i(x) = V(x) + \lambda^T g(x), \quad (92)$$

де $x \in X$, $g(x) = [g_1(x), \dots, g_m(x)]^T$, $\lambda \in \Lambda_0 = \{\lambda_1, \dots, \lambda_m\}^T$.

Сідловою точкою функції Лагранжа називається така пара $(x^*, \lambda^*) \in X_0 \times \Lambda_0$, що за усіх $x \in X_0$, $\lambda \in \Lambda_0$

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*). \quad (93)$$

Співвідношення (93) записується також як

$$L(x^*, \lambda^*) = \min_{x \in X_0} \max_{\lambda \in \Lambda_0} L(x, \lambda) = \max_{\lambda \in \Lambda_0} \min_{x \in X_0} L(x, \lambda).$$

Вирішення різноманітних задач медико-біологічних досліджень пов'язане із оптимізацією, тобто відшукуванням таких значень факторів $x_i = x_i^*$, $i = \overline{1, k}$, за яких забезпечується екстремальне (мінімальне чи максимальне) значення вихідного показника $y^* = f(x_1^*, \dots, x_k^*)$.

Методи, що застосовуються для пошуку екстремуму, називають методами *екстремального експерименту*. Вони передбачають дослідження ділянок поверхні відгуку $y = f(x_1, \dots, x_k)$ шляхом постановки спланованих серій дослідів, кожна із яких надає можливість отримати інформацію про напрямок руху до екстремуму. Отже, екстремальне значення відгуку досягається за допомогою многократної послідовної процедури вивчення поверхні відгуку і просування у факторному просторі.

Існують різні методи оптимізації (градієнтні, методи випадкового пошуку тощо),

що відрізняються способами визначення руху та організації саме руху до екстремуму.

У звичайній практиці експериментатор часто використовує метод почергової оптимізації за факторами, відомий як метод Гауса--Зейделя . У такій "класичній" схемі оптимізації послідовно відшукують часткові екстремуми функції відгуку за кожним фактором, за умови фіксованих решти факторів. Очевидна громіздкість цього методу, особливо за великої кількості факторів. За складних форм поверхні відгуку такий метод може призвести до помилкових рішень.

Тому використовують ефективніші методи, зокрема, метод крутого сходження та симплексний метод оптимізації .

Метод крутого сходження

Цей метод запропонували П. Бокс і В. Вільсон (1951). Він передбачає "покрокове" вивчення поверхні відгуку для визначення напрямку і подальшого просування до оптимуму. Стратегія пошуку екстремуму полягає у послідовній постановці невеликої серії дослідів із використанням ПФЕ або ДФЕ типу 2^n із подальшим рухом у напрямку градієнту функції відгуку. Такий напрямок відповідає найкоротшому шляху до оптимуму. На першому етапі в околі деякого початкового режиму реалізують план типу 2^k , за результатами якого будують лінійне рівняння регресії і виконують його статистичний аналіз (перевіряють значимість коефіцієнтів b_i та адекватність рівняння). Адекватність рівняння надає можливість перейти до руху у напрямку градієнту функції відгуку:

$$\nabla y = \frac{\partial y}{\partial x_1} i + \frac{\partial y}{\partial x_2} j + \dots + \frac{\partial y}{\partial x_l} l,$$

де $i, j, \dots, l$ -- одиничні вектори у напрямку координатних осей.

Для здійснення указанного руху необхідно змінювати фактори пропорційно отриманим коефіцієнтам b_i із урахуванням їхнього знаку. Оскільки рух у напрямку градієнту передбачає вихід за межі області, що вивчена у факторному експерименті, метод крутого сходження є використання лінійного рівняння регресії для екстраполяції розрахунків відгуку у .

Рух за градієнтом реалізують доти, доки спостерігається покращення значень параметру оптимізації. Якщо у першій серії дослідів досягти оптимуму не вдається, в околі кращого відшуканого режиму ставлять новий експеримент 2^k , визначають новий напрямок руху за градієнтом тощо. Така процедура постановки планів 2^k із подальшим рухом за градієнтом виконується до досягнення області розташування оптимального режиму -- стаціонарної області. Дослідження цієї області здійснюють за допомогою складніших планів, що надають можливість отримати її опис у вигляді нелінійних функцій (зазвичай, квадратичних). Математичний аналіз таких функцій дозволяє отримати координати екстремуму і характер поверхні відгуку у цьому околі.

Практична процедура пошуку екстремуму за методом крутого сходження здійснюється у такій послідовності.

1. В околі вхідного режиму, що відповідає кращим відомим умовам, проводиться ПФЕ чи ДФЕ типу 2^k . Вибір початкової точки, кроків варіювання за факторами і подання даних у вигляді матриці планування для кодованих значень факторів виконується як звичайно. Особливість полягає лише у тому, що кроки варіювання вибираються не занадто великими, щоб ділянка поверхні відгуку, що досліджується, не значно відрізнялася від лінійної і її можна було адекватно описати поліномом першого степеню. Далі отримані результати експерименту 2^k підпорядковуються звичайній процедурі

статистичного аналізу:

- а) перевірку відновлюваності експерименту та обчислення похибки досліду s_y ;
- б) перевірку значимості коефіцієнтів b_i лінійної моделі;
- б) перевірку адекватності отриманої лінійної моделі:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k.$$

2. Обчислення добутку коефіцієнтів на кроки варіювання: $b_i \times \lambda_i$. Фактор із максимальним значенням $b_\delta \times \lambda_\delta$ приймається за базовий, тобто для цього фактору

$$(b_\delta \times \lambda_\delta) = \max(b_i \times \lambda_i).$$

3. Для базового фактору вибирають крок для руху за крутим сходженням λ'_δ . Зазвичай залишають старий крок $\lambda'_\delta = \lambda_\delta$ або менший для руху у напрямку оптимуму.

4. Визначають величини кроків для інших факторів λ'_i . Оскільки для руху за градієнтом фактори мають змінюватися пропорційно відповідним коефіцієнтам нахилу b_i , λ'_i відшукують шляхом множення коефіцієнтів b_i на відношення добутків:

$$I_i = \frac{b_i \cdot \lambda_i}{b_\delta \cdot \lambda_\delta},$$

Тобто

$$\lambda'_i = I_i \lambda'_\delta.$$

5. Розраховують умови дослідів за крутим сходженням шляхом одночасного додавання або віднімання (залежно від знаку b_i) кроків до початкового рівня фактору x_{i0} . "Передбачені" значення відгуку у вказаній серії "уявних" дослідів визначають за формулою:

$$\hat{y} = b_0 + h \sum_{i=1}^k b_i \lambda'_i,$$

де $h = 1, 2, 3, \dots$ -- номер кроку. Такий розрахунок можна виконувати до m -го кроку, тобто до досліду, в якому $y_k = y_m$, y_m -- максимально можливий вихід (наприклад, вихід продукту у відсотках не може перевищити 100%).

6. Реалізують деякі із уявних дослідів для перевірки ефективності руху за крутим сходженням і відповідності передбачених значень відгуку значенням, що спостерігаються в експерименті. Реалізують дві--три точки траєкторії і оцінюють їхні результати.

7. Умови кращого досліду, що отримано за крутого сходження, приймаються за основний рівень факторів у наступній серії дослідів. В околі цього досліду реалізують новий експеримент, визначають нове круте сходження тощо. Оскільки під час просування до оптимуму із кожним циклом крутого сходження кривизна поверхні відгуку зростає, для отримання адекватного лінійного наближення необхідно зменшувати кроки λ_i і λ'_i після кожного наступного циклу.

7. При досягненні стаціонарної області усі коефіцієнти b_i лінійної моделі стають незначимими. На цьому експеримент зупиняється. Іншим рішенням є побудова лінійного плану до плану другого порядку з метою детального дослідження області оптимуму.

В якості **прикладу** наведемо розрахунок крутого сходження в задачі дослідження умов електричної стимуляції діяльності шлунку собак.

Мета роботи полягала у відшукуванні області оптимальних режимів електростимуляції, які б забезпечили повне відновлення ослабленої активності шлунку

собак за допомогою макету електростимулятора, що надає можливість подавати імпульси визначеної протяжності та амплітуди синхронно із біоелектричним ритмом шлунку. Ослаблення перистальтичної діяльності шлунку досягалося шляхом впровадження 1--3 мл 5% розчину гексеналу.

Основними факторами електростимуляції були: протяжність імпульсу, що стимулює, X_1 (мс), амплітуда імпульсу, що стимулює, X_2 (мВ), момент подачі імпульсу, що стимулює, що задається напругою виконання пристрою, що стимулює, X_3 (мВ). Ефект стимуляції визначався величиною амплітуди біопотенціалів шлунку y (мВ), що характеризує силу скорочень гладких мускулів.

На першому етапі дослідження для визначення зони чутливості електричної стимуляції, тобто області значень факторів X_1, X_2 , які викликають значимі змінювання амплітуди y , реалізовано план 2^2 . Центр експерименту відповідав умовам, близьким до оптимального режиму стимулювання для моторики шлунку білих пацюків. Результати експерименту 2^2 і дослідів крутого сходження в область зони чутливості електростимуляції при $X_3 = 0$ наведено у таблиці

Ошибка! Источник ссылки не найден..

Після перевірки однорідності дисперсій за критерієм Кохрена, яка дала позитивні результати, за даними експерименту 2^2 обчислено коефіцієнти регресії:

$$b_0 = \frac{1}{4}(+0,6 + 0,7 + 0,7 + 0,8) = 0,7, \quad b_1 = \frac{1}{4}(-0,6 - 0,7 + 0,7 + 0,8) = 0,05,$$

$$b_2 = \frac{1}{4}(-0,6 + 0,7 - 0,7 + 0,8) = 0,05, \quad b_{12} = 0.$$

Похибка дослідів визначалася за даними паралельних вимірювань на 3-х тваринах:

$$s_y^2 = \frac{1}{4(3-1)} \sum_{k=1}^4 \sum_{i=1}^3 (y_{ik} - \bar{y}_k)^2 = 0,110^2, \quad s_{\bar{y}} = \frac{s_y}{\sqrt{3}} = 0,038;$$

$$s(b) = \frac{s_{\bar{y}}}{\sqrt{4}} = 0,019, \quad f = 8.$$

Значення b_1, b_2 виявилися більші за відповідний довірчий інтервал

$$\Delta b = t_{0,05} s(b) = 2,3 \cdot 0,019 = 0,045.$$

Лінійна модель адекватно описувала дані експерименту:

$$F = \frac{s_{\text{аа}}^2}{s_{\bar{y}}^2} = \frac{0}{0,038^2} = 0 < F_{0,05}(1,8) = 5,3.$$

Після обчислення добутків $b_1 \times \lambda_1 = 0,05 \cdot 100 = 5$ і $b_2 \times \lambda_2 = 0,05 \cdot 70 = 3,5$ для базового фактору X_1 із максимальним добутком вибрано крок для крутого сходження $\lambda'_1 = 80$, значення кроку для X_2 : $\lambda'_2 = 3,5 \cdot 80 = 56 \approx 60$.

Після додавання до початкових рівнів $X_{10} = 200$, $X_{20} = 140$ зазначених кроків розраховано умови для 5-ти дослідів крутого сходження.

У дослідях 4 і 5 рух виконувався тільки по X_2 за значення $X_1 = 440$ внаслідок наявних обмежень. Передбачені значення відгуку у цих точках

$$y = b_0 + h \sum_i b_i \cdot \lambda'_i, \quad h = 1, 2,$$

$$y_k = 0,7 + [0,05 \frac{80}{100} + 0,05 \frac{60}{70}]h = 0,7 + 0,08h = [0,78; 0,86]$$

тощо.

Реалізація вказаних дослідів засвідчила, що збільшення y зі зростанням X_1, X_2 здійснюється швидше за передбачення за лінійним рівнянням. Зокрема, у дослідах 4 і 5 спостерігалось повне відновлення рухливої активності шлунку до початкової фізіологічної норми із $y = 2$. Тому подальший рух за крутим сходженням не являє цікавості і був спланований експеримент другого порядку із урахуванням фактору X_3 -моменту подачі імпульсу, що стимулює, з метою отримання адекватної моделі для опису цієї області.

Табл.4.10. Матриця плану і результати дослідів

Фактори	x_1	x_2	y	$\hat{y}$
Рівні				
0	200	140	0,7	
(-)	100	70		
(+)	300	210		
Досліди				
1	-	-	0,6	
2	-	+	0,7	
3	+	-	0,7	
4	+	+	0,9	
b	0,05	0,05		
$b_i \cdot \lambda_i$	5	3,5		
λ'_i	80	80		
Досліди К.С.				
1	280	200	0,8	0,78
2	300	260	1,0	0,86
3	440	320	1,2	0,94
4	440	380	1,9	0,98
5	440	440	2,2	1,02

4.12 Симплексний метод оптимізації

Симплексний метод оптимізації призначений для математичного опису властивостей багатокомпонентних систем і сумішей, коли на фактор накладено обмеження вигляду

$$\sum_{i=1}^k x_i = 1,$$

де x_i -- вміст i -го компоненту у суміші. Формулювання таких задач можливе, наприклад, у фармакології при вивченні властивостей ліків залежно від компонентного складу; у задачах комплексної терапії при дослідженні корисності комбінованої дії кількох ліків з метою відшукування терапевтичного синергізму тощо.

Симплексом називають випуклу фігуру, що утворена $k+1$ вершиною у просторі k факторів, причому ці вершини не належать одночасно жодному із підпросторів із $k-1$

факторів. Симплекс називають регулярним, якщо усі відстані між його вершинами є рівні.

Симплексний метод надає можливість суміщувати пробні досліди для визначення напрямку руху із робочим рухом по поверхні відгуку до області оптимуму. Якщо в усіх $k+1$ вершинах симплексу поставити досліди і виміряти відгук, за величиною відгуку у вершинах можна визначити напрямок руху для наближення до екстремуму. Після проведення серії дослідів, що поставлено у вершинах правильного симплексу, визначається точка, що відповідає умовам, за яких отримуються найгірші результати.

Потім використовується важлива властивість симплексу, за якою із будь-якого симплексу шляхом відкидання однієї із вершин можна отримати новий симплекс, замінюючи відкинуту вершину її дзеркальним відображенням відносно протилежної його грані.

Якщо відкинути точку із найгіршими результатами і побудувати на грані, що залишилася новий симплекс, його центр зміститься у напрямку: найгірша точка -- центр тяжіння решти точок, тобто у напрямку до екстремуму. Потім процес відкидання вершини із найгіршим значенням цільової функції і побудови нового симплексу повторюється.

У результаті утворюється ланцюг симплексів, що рухаються у факторному просторі до точки екстремуму. Отже, рух до екстремуму здійснюється шляхом дзеркального відображення точки із найгіршими результатами відносно центру протилежної грані симплексу.

Порядок дій при використанні симплексного методу такий:

1. Обирають початкову точку C_1 , а також інтервали варіювання Δx_i усіх факторів ($i = \overline{1, k}$).

2. Обирають безрозмірну величину ρ_{sim} сторони (або ребра) симплексу у відносних одиницях відносно до інтервалів варіювання Δx_i . Найпростіше обирати $\rho_{sym} = 1$. Прагнуть щоб у безрозмірних одиницях сторони симплексу були рівними.

3. Обчислюють координати решти вершин початкового симплексу. Для цього через початкову точку C_1 проводять осьові лінії, паралельні координатним осям і вибирають квадрант, в якому за припущенням, має розташовуватися екстремум цільової функції. У початкову точку розміщують вершину симплексу C_1 , а сам симплекс (рис. 4.2) розташовують так, щоб його сторони утворили із осьовими лініями рівні кути.

За такого розташування початкового симплексу координати його вершин визначають за допомогою матриці (рис. 4.3), в якій дано координати вершин $(k+1)$ -вимірного симплексу в n -факторному просторі.

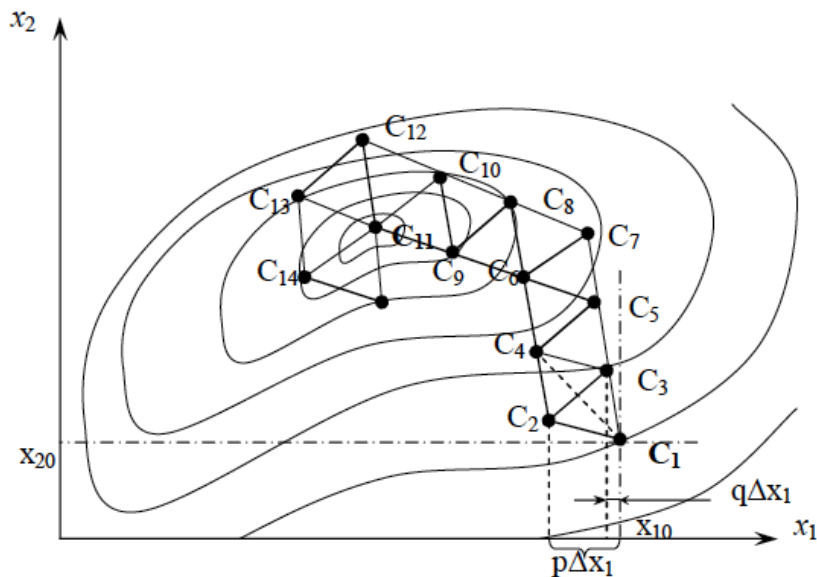


Рис.4.2 Пошук екстремуму функції відгуку симплексним методом

Безрозмірні відносні величини p і q за такого розташування симплексу визначають за формулою

$$p = (\sqrt{k+1} + k - 1)/(k\sqrt{2}), \quad q = (\sqrt{k+1} - 1)/(k\sqrt{2}).$$

На рис. **Ошибка! Источник ссылки не найден.** показані розміри $p\Delta x_i$ і $q\Delta x_i$ при $\rho_{sim} = 1$. Якщо $\rho_{sim} \neq 1$, Δx_i помножують на ρ_{sim} . Знаки Δx_i залежать від номера квадранту, в якому розташований початковий симплекс. Для $k = 2$ маємо $p \approx 0,966$, $q \approx 0,259$.

На відміну від методу крутого сходження, де спочатку реалізують плани 2^k (пробні кроки) і тільки потім за отриманою лінійною моделлю обчислюють напрямки градієнту і здійснюють рух до оптимуму (робочі кроки), у цьому методі пробні і робочі кроки суміщені.

Факторы x_i	x_1	x_2	x_3	...	x_i	...	x_k
Вершина C_1	x_{10}	x_{20}	x_{30}	...	x_{i0}	...	x_{k0}
Вершина C_2	$x_{10} + p\Delta x_1$	$x_{20} + q\Delta x_2$	$x_{30} + q\Delta x_3$	...	$x_{i0} + q\Delta x_i$	...	$x_{k0} + q\Delta x_k$
Вершина C_3	$x_{10} + q\Delta x_1$	$x_{20} + p\Delta x_2$	$x_{30} + q\Delta x_3$	...	$x_{i0} + q\Delta x_i$	...	$x_{k0} + q\Delta x_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
Вершина C_{i+1}	$x_{10} + q\Delta x_1$	$x_{20} + q\Delta x_2$	$x_{30} + q\Delta x_3$	...	$x_{i0} + p\Delta x_i$	...	$x_{k0} + q\Delta x_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
Вершина C_{k+1}	$x_{10} + q\Delta x_1$	$x_{20} + q\Delta x_2$	$x_{30} + q\Delta x_3$	...	$x_{i0} + q\Delta x_i$	...	$x_{k0} + p\Delta x_k$

Рис.4.3 Завдання координат вершин симплексу

4. У вершинах симплексу виконують спостереження відгуку і порівнюють за величиною; вибирають вершину із мінімальним відгуком і відображають її відносно протилежної сторони або грані; відшукують вершину наступного симплексу II, n вершин якого одночасно є і вершини попереднього симплексу I. Координати вершини, що

відображається, обчислюють за формулою

$$x_{il,m+1} = [2(x_{i1,m} + x_{i2,m} + \dots + x_{i,n+1,m})/k] \pm x_{il,m},$$

де i -- номер фактора ($i = \overline{1, k}$); l -- номер вершини m -го симплексу, де виявлено мінімальний (у разі пошуку максимуму) відгук;

$m+1$ -- номер наступного симплексу, який містить вершину, що відображається (їй умовно привласнюють той самий номер l); k -- число факторів.

Якщо мінімальний відгук виявився одразу у двох вершинах, питання яку із них відображати, вирішують довільно.

5. Ставлять експеримент у вершині, що відображена, нового симплексу і відгук у цій вершині порівнюють із відгуками у решті вершин, а потім знову вибирають вершину із мінімальним відгуком і відображають її через протилежну сторону (або грань) симплексу. Якщо у новій вершині ($m+1$) симплексу відгук виявився знову мінімальним, повертаються до m -го симплексу і відображають другу за мінімальністю вершину тощо.

6. Експеримент триває доти, доки симплекс не здійснить повний оберт навколо однієї з вершин. На рис. **Ошибка! Источник ссылки не найден.** це вершина C_{11} .

Для завдання експериментальної області можуть використовуватися геометричні фігури, відмінні від квадрату, кубу та гіперкубу, що використовуються в ортогональних планах 2^k . У симплекс-плануванні розташування дослідів здійснюється у вершинах регулярного (правильного) симплексу.

Процедуру побудови матриці симплекс-плану можна пояснити на прикладі двох незалежних змінних x_1, x_2 . Рівнобічний трикутник у площині кодованих змінних можна розташувати так, щоб для вектор-стовпців, елементами яких є координати вершин, виконувалися умови симетричності і ортогональності:

$$\begin{bmatrix} a & b \\ -a & b \\ 0 & -2b \end{bmatrix}.$$

Значення a і b можна обчислити із умови нормування

$$\sum_{i=1}^N x_{i1}^2 = \sum_{i=1}^N x_{i2}^2 = N,$$

за якого оцінки коефіцієнтів b_j лінійної моделі матимуть однакові дисперсії. Із урахуванням цієї умови для стовпців матриці отримаємо $2a^2 = 6b^2 = 3$, отже, матрицю незалежних змінних x_1, x_2 :

$$\begin{bmatrix} 1,288 & 0,707 \\ -1,288 & 0,707 \\ 0 & -1,414 \end{bmatrix}.$$

Для зручності виконаємо масштабування елементів матриці, щоб отримати правильний симплекс, що вписаний у коло одиничного радіусу. Після ділення усіх елементів на 1,414 отримуємо матрицю симплекс-планування:

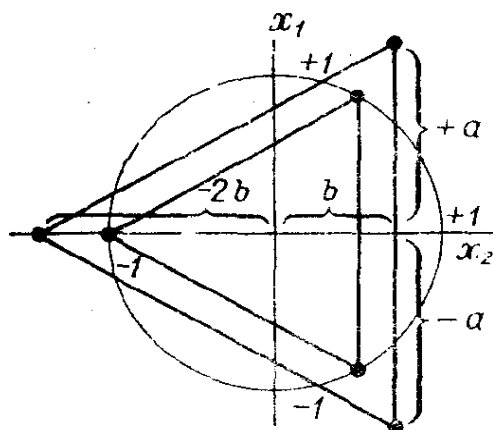


Рис.4.4 Геометрична інтерпретація симплексу

Табл.4.11. Матриця симплекс-планування

n	x_0	x_1	x_2	y
1	+1	+0,866	0,500	y_1
2	+1	-0,866	0,500	y_2
3	+1	0	-1	y_3

Геометрична інтерпретація симплексу наведена на рис. 4.4.

Для чотирьох змінних вихідна матриця симплексу матиме вигляд:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ -a_1 & a_2 & a_3 & a_4 \\ 0 & -2a_2 & a_3 & a_4 \\ 0 & 0 & -3a_3 & a_4 \\ 0 & 0 & 0 & -4a_4 \end{bmatrix}.$$

Із умови нормування отримаємо $2a_1^2 = 6a_2^2 = 12a_3^2 = 20a_4^2 = 5$, звідки

$$a_1 = 1,580, a_2 = 0,912, a_3 = 0,644, a_4 = 0,500.$$

Після ділення усіх елементів на 2,00 отримаємо матрицю планування:

Табл.4.13.Матриця симплекс-планування

n	x_0	x_1	x_2	x_3	x_4	y
1	+1	+0,790	0,456	0,322	0,250	y_1
2	+1	-0,790	0,456	0,322	0,250	y_2
3	+1	0	-0,912	0,322	0,250	y_3
4	+1	0	0	-0,966	0,250	y_4
5	+1	0	0	0	-1,000	y_5

Матриця центрованого K -симплексу, вписаного у сферу одиничного радіусу, може бути записана у загальному вигляді як Для чотирьох змінних вихідна матриця

симплексу матиме вигляд:

$$\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_k \\ -a_1 & a_2 & a_3 & \dots & a_k \\ 0 & -2a_2 & a_3 & \dots & a_k \\ 0 & 0 & -3a_3 & \dots & a_k \\ \dots & & & & \\ 0 & 0 & 0 & \dots & -ka_k \end{bmatrix}.$$

де

$$a_i = \sqrt{\frac{k+1}{ik(i+1)}}, \quad i \leq k.$$

В силу властивості ортогональності, симетричності і нормування матриці симплекс-плану обчислення коефіцієнтів лінійної регресійної моделі, аналіз її придатності для опису результатів експерименту виконуються за аналогією із повним факторним експериментом.

Дисперсії оцінок коефіцієнтів у симплекс-планах та планах 2^k однакові і визначаються як

$$\sigma^2[b_j] = \frac{\sigma^2[y]}{\sum_{i=1}^N s_{ij}^2} = \frac{\sigma^2[y]}{N}.$$

Проте симплекс-плани мають більшу дисперсію передбачених значень залежної змінної y . Так, наприклад, для лінійної моделі плану 2^{4-1}

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

дорівнює

$$\sigma^2[y] = \sigma^2[b_0] + x_1^2\sigma^2[b_1] + x_2^2\sigma^2[b_2] + x_3^2\sigma^2[b_3] + x_4^2\sigma^2[b_4].$$

Із урахуванням масштабного зменшення у два рази усіх елементів матриці симплекс-плану лінійну модель слід записувати у вигляді

$$y = b_0 + 2b_1x_1 + 2b_2x_2 + 2b_3x_3 + 2b_4x_4$$

$$\sigma^2[y] = \sigma^2[b_0] + 4x_1^2\sigma^2[b_1] + 4x_2^2\sigma^2[b_2] + 4x_3^2\sigma^2[b_3] + 4x_4^2\sigma^2[b_4].$$

Максимальні значення $\sigma^2[\hat{y}]$ досягаються на межі області експерименту.

На ґрунті правильних симплексів можна побудувати досить формалізовану процедуру руху до екстремуму.

Симплекс-процедура руху полягає у таких етапах, що повторюються:

1. Реалізація дослідів відповідно до обраної матриці симплекс-плану;
2. Відкидання вершини симплексу із мінімальним значенням вихідної змінної (параметру оптимізації);
3. Побудова на грані, що залишилася, нового симплексу, в якого відкинута вершина замінюється її дзеркальним відображенням. Рух у просторі вхідних змінних, наприклад, у напрямку максимуму функції відгуку, здійснюється шляхом дзеркального відображення вершини X_1 симплексу із найменшим значенням X_1 і побудові нового симплексу.

Координати нової вершини симплексу визначаються із рівностей:

$$x_{ij}^* = \frac{2}{k} (x_{1j} + x_{2j} + \dots + x_{i-1,j}x_{i+1,j} + \dots + x_{k+1,j}) - x_{ij},$$

де x_{ij}^* -- j -та координата нової вершини; $j = \overline{1, k}$ -- число вхідних змінних; $i = \overline{1, N}$ -- число вершин симплексу; $N = k + 1$.

При реалізації дослідів згідно симплекс-плану отримані у всіх його вершинах результати порівнюють, відкидають вершину із найгіршим результатом і ставлять новий експеримент у точці, що є дзеркальна до відкинутої. Так, шляхом послідовного відкидання вершин здійснюється переміщення симплексу із кожним дослідом у напрямку до оптимуму. Центр симплексу при цьому описуватиме траєкторію, що є близька до траєкторії руху за градієнтом у методі крутого сходження. Процедура оптимізації для двох змінних за симплекс-методом наведена на рис.

Ошибки! Источник ссылки не найден.. У початковому симплексі 123, що реалізовано в околі режиму із X_{10} і X_{20} , відкидають вершину 2 із найгіршим результатом y_2 , потім у наступних трикутниках послідовно відкидають вершини 1, 3, 4 тощо із гіршими результатами y_1 , y_2 , y_4 тощо. Ламана лінія показує рух центру симплексу у процесі оптимізації. При симплекс-процедурі не потрібно точно вимірювати значення вихідного показника, досить їх проранжувати, щоб розрізнити гірші значення. За допомогою симплексної процедури можна ефективно вирішувати задачі оптимізації об'єктів, що дрейфують (наприклад, об'єктів біотехнології). У цьому випадку симплекс відслідкуватиме оптимум, що дрейфує.

Із-за похибок, що викликаються некерованими факторами, а також наявністю гребенів і ярів на поверхні відгуку рух симплексу стає нестійким. Симплекс може зациклюватися біля вершини із завищеним за рахунок похибок відгуком або біля грані, що опинилася на гребні. Але деяким ускладненням логіки руху можна запобігти цій проблемі. Якщо симплекс почав обертатися навколо однієї із вершин [ця вершина трапляється більше за $k+1$) послідовних симплексах], для прояснення ситуації треба повторити експеримент у цій вершині. Якщо отримано високий результат у цій вершині, приймається рішення про досягнення оптимуму. Інакше високе значення відгуку, що отримано на попередньому циклі, вважається помилковим, і ця точка у подальшому відкидається.

Якщо симплекс почав обертатися навколо грані, тобто у новому симплексі точка виявилася найгіршою і здійснюється повернення до початкового симплексу, треба відкинути наступну за порядком малості відгуку вершину.

Під час руху симплексу у процесі оптимізації алгоритм обчислень зводиться до визначення координат нової точки

$$x_{ji}^* = \frac{2}{k}(x_{1i} + x_{2i} + \dots + x_{ki}) - x_{ji},$$

де j -- номер вершини, що відкидається, із найгіршим значенням відгуку; i -- індекс координати ($i = \overline{1, k}$).

Основні переваги симплексного методу оптимізації:

1. Метод сповна формалізований і надає чітке правило, коли і як треба змінювати умови процесу.
2. Використання методу не потребує виконання статистичного аналізу результатів і не пред'являє строгих вимог до адекватності лінійного наближення області, що досліджується, поверхні відгуку.
3. У процесі оптимізації обчислення прості і зводяться на кожному кроці до визначення координат однієї нової точки.
4. Для прийняття рішень про постановку нового дослідів не потрібно точно

вимірювати значення вихідного показника, досить їх ранжувати, щоб розрізнити гірші значення. Ця перевага робить метод придатним у задачах оптимізації, коли вихід оцінюється якісно, а також у випадку великої кількості відгуків або наявності обмежень на значення факторів $x'_i \leq x_i \leq x''_i$ і відгуків $y'_m \leq y_m \leq y''_m$.

5. На будь-якому етапі оптимізації легко додати ще один фактор шляхом додатково запровадження у симплекс тільки однієї точки, яка разом із іншими утворює симплекс із розмірністю на одиницю більше.

6. У процесі оптимізації рух до оптимуму здійснюється після кожного дослідів, а не після серії дослідів.

7. Метод добре пристосований для оптимізації об'єктів, що дрейфують. У цьому разі симплекс відслідковує оптимум, що дрейфує.

7. Постановка дослідів не пред'являє високих вимог до регулярності симплексу, тобто строгій підтримці факторів на заданих рівнях.

Недоліки симплексного методу оптимізації.

1. Метод надає обмежену інформацію про поверхню відгуку і не оцінює взаємозв'язок факторів.

2. Усі фактори мають бути кількісними.

3. Ефективність методу спадає зі зростанням похибок експерименту. При цьому здійснюється більше помилкових рухів на шляху до оптимуму із зациклюванням в окремих точках.

Для запобігання останнього недоліку використовують модифікацію симплекс - методу, в якому розмір симплексу регулюється залежно від рівня шумів -- похибки експерименту.

Симплекс-метод слід використовувати тоді, коли результати дослідів можна отримати за короткий час (наприклад, коли відгуком є скорочення м'язового пучка, змінювання кров'яного тиску, показників діяльності шлунку, серця та інших органів у задачах їхньої стимуляції тощо).

У якості прикладу розглянемо ту саму задачу підбору умов електростимуляції шлунку собак (протяжності та амплітуди імпульсів, що стимулюють, X_1, X_2).

При інтервалах варіювання $\Delta X_1 = 100$ мс, $\Delta X_2 = 70$ мВ розмір симплексу -- відстань між двома вершинами (в одиницях варіювання відповідних факторів) було прийнято рівним $d = 2$. У цьому разі висота симплексу дорівнює сумі радіусів R_2 і r_2 кіл, що описані біля нього і вписані у нього, із множителем $d = 2$; $d \cdot (r_2 + R_2) = 2 \cdot (0,289 + 0,578) = 1,72$.

Вочевидь, координати вершин вхідного симплексу у кодованих і натуральних значеннях факторів будуть

$$x_{1A} = 0, x_{2A} = 0, X_{1A} = 100, X_{2A} = 70;$$

$$x_{1B} = 0,86, x_{2B} = 0,5, X_{1B} = 100 + 2 \cdot 100 \cdot 0,86 = 272, X_{2B} = 70 + 2 \cdot 0,5 \cdot 70 = 140;$$

$$x_{1C} = 0, x_{2C} = 1, X_{1C} = 100 + 2 \cdot 100 \cdot 100, X_{2C} = 70 + 2 \cdot 1 \cdot 70 = 210.$$

При реалізації вхідного симплексу найменшим виявився відгук у вершині А: $y_A = 0,6$. Будуємо її дзеркальне відображення, тобто визначаємо координати нової точки A^* у кодованих і натуральних значеннях факторів:

$$x_{1A}^* = \frac{2}{2}(x_{1B} + x_{1C}) - x_{1A} = 0,86 + 0 - 0 = 0,86, X_{1A}^* = 100 + 2 \cdot 100 \cdot 0,86 = 272;$$

$$x_{2A}^* = \frac{2}{2}(x_{2B} + x_{2C}) - x_{2A} = 0,5 + 1 - 0 = 1,5, \quad X_{2A}^* = 70 + 2 \cdot 70 \cdot 1,5 = 280.$$

У новому симплексі A^*BC найменше значення відгуку виявилось у точці C : $y_C = 0,7$. Дзеркальне відображення цієї точки має координати:

$$x_{1C}^* = \frac{2}{2}(x_{1B} + x_{1A}^*) - x_{1C} = 0,86 + 0,86 - 0 = 1,72,$$

$$x_{2C}^* = \frac{2}{2}(x_{2B} + x_{2A}^*) - x_{2C} = 0,5 + 1,5 - 1 = 1,0.$$

4.11.1 Симплексний метод

Сутність методу полягає у тому, що рух до максимуму функції η , що задана на $X = R^n$, здійснюється послідовним відображенням вершин симплексу. За визначенням n -вимірний симплекс являє собою многогранник, що утворений $n+1$ точками (вершинами), що не належать одночасно жодному підпростору меншої розмірності. Симплекс зветься *регулярним*, якщо відстані між його вершинами рівні.

Найпростішим варіантом симплексного методу є так званий *послідовний симплексний метод*, що використовує дзеркальне відображення регулярних симплексів відносно граней, що є протилежні вершинам, в яких значення випадкової величини y є найменше. Многочасне відображення симплексу призводить до крокового руху центру симплексу до мети за траєкторією деякої лінії.

Запровадимо матрицю розміром $n \times (n+1)$:

$$A = \begin{bmatrix} 0 & 0 & \dots & 0 \\ p_n & q_n & \dots & q_n \\ q_n & p_n & \dots & q_n \\ \dots & \dots & \dots & \dots \\ q_n & q_n & \dots & p_n \end{bmatrix}.$$

$$\text{Тут } q_n = \frac{1}{n\sqrt{2}}(\sqrt{n+1}-1), \quad p_n = \frac{1}{n\sqrt{2}}(\sqrt{n+1}+n-1).$$

Координати z_{ij} ($i = \overline{1, n}, j = \overline{1, n+1}$) вершин z_j регулярного симплексу з довжиною ребра $L = 1$ визначаються рядками матриці A . При цьому вершина z_1 буде міститися у початку координат, а вектори, що відповідають вершинам $z_2, \dots, z_{n+1}$, складуть однакові кути із координатними осями.

Координати вершин регулярного симплексу із довжиною ребра $L = 1$ і центром у початку координат визначаються рядками матриці

$$\begin{bmatrix} -r_1 & -r_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ R_1 & -r_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ 0 & R_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix},$$

де $r_i = 1/\sqrt{2i(i+1)}$; $R_i = \sqrt{i/(2(i+1))}$, $i = \overline{1, n}$.

Величини r_i і R_i -- радіуси вписаної і описаної сфер для i -вимірного регулярного симплексу із довжиною ребра $L = 1$.

Якщо треба побудувати регулярний симплекс із довжиною ребра $\neq 1$ і центром (вершиною) у точці x_0 , що відрізняється від початку координат, до координат вершин симплексу додаються відповідні координати точки x_0 , а потім отримані величини множаться на L .

Алгоритм послідовного симплексного пошуку полягає у такому:

- 1) Будуємо регулярний симплекс Σ_0 , покладаємо $k = 1$.
- 2) Виконуємо вимірювання у вершинах симплексу Σ_0 .
- 3) Вибираємо найменшу величину y_l із результатів вимірювання $y_1, \dots, y_{n+1}$ у вершинах симплексу Σ_{k-1} .
- 4) Будуємо симплекс Σ_k шляхом заміни точки z_l в якій отримано значення y_l новою точкою z'_l , координати якої дорівнюють

$$z'_{il} = \frac{2}{n} \sum_{j=1}^{n+1} z_{ij} - (1 + 2/n) z_{il}, \quad i = \overline{1, n+1}$$

(точка z'_l із координатами z'_{il} є дзеркальне відображення точки z_l відносно протилежної грані симплексу Σ_{k-1}).

- 5) Виконуємо вимірювання у точці z'_l , отримуємо y'_l .
- 6) Якщо $y'_l < y_i$ ($i = \overline{1, n+1}$, $i \neq l$), в якості z_l вибираємо вершину симплексу Σ_{k-1} , яка відповідає другому за величиною значенню із набору $\{y_1, \dots, y_{n+1}\}$, і переходимо до кроку 4.
- 7) Якщо $k > n+1$ і на попередніх $n+1$ ітераціях найбільше значення у наборі $\{y_1, \dots, y_{n+1}\}$ залишалося одним і тим самим, повторюємо вимірювання у точці, що відповідає найбільшому значенню.
- 8) Замінюємо k на $k+1$, y_l на y'_l і переходимо на крок 3.

За допомогою цього методу вдається досить швидко відшукувати область максимуму (і відслідковувати її, якщо функція η нестационарна) при вирішенні багатьох задач, в яких вимірювання змінних, що контролюються, $x \in X$ може виконуватися із похибкою, а випадкові величини y , що обчислено у різних точках x , можуть бути залежними між собою і залежати від часу.

Обмеження типу нерівностей при проведенні симплексного пошуку враховуються дуже просто: вершини, що не відповідають обмеженням, відкидаються.

Сталий розмір симплексу не забезпечує одночасно високу швидкість руху симплексу у початку пошуку і точність відшукання екстремуму у його кінці. Тому зазвичай розміри симплексу зменшують зі зростанням номера кроку.

Наведемо одну із типових модифікацій послідовного симплексного методу, в якій розмір симплексу на кожному кроці зменшується.

Після відображення симплексу на кожному кроці вибирається вершина, в якій випадкова величина y набула найбільше значення, початок координат переноситься у цю точку, після чого координати решти вершин перетворюються за формулою

$$z'_{ij} = L_k z_{ij}, \quad i = \overline{1, n}, \quad j = \overline{1, n+1}.$$

k -- номер кроку; $\{L_k\}$ -- послідовність чисел, яка визначає закон змінювання ребер симплексу. Оскільки зазвичай вибирають $L_k \approx 1$ і припускають, що функція η змінюється досить повільно, значення випадкової величини y для вершин нового симплексу не обчислюють, а використовують значення цієї випадкової величини y у відповідних вершинах старого симплексу. Виняток складають ті вершини нового симплексу, в яких упродовж $n+1$ попередніх кроків значення y не обчислювалося; у цих вершинах обчислюються значення y .

Отже, на кожному кроці, крім нульового, потрібно у середньому біля одного обчислення значень y . Послідовність $\{L_k\}$ і довжина ребра початкового симплексу вибираються із апріорних міркувань про бажану швидкість руху симплексу на початковому і кінцевому етапах процесу.

Симплексний пошук має властивість, що змінювання довжини кроку y напрямку псевдоградієнту γ_k однозначно пов'язане зі змінюванням інтервалу варіювання, або "довжини пробного кроку" α_k .

4.11.2 Симплекс-метод оптимального планування експерименту

Суть методу полягає у тому, що многофакторний план представляється у вигляді опуклого многогранника; вершини многогранника відповідають значенням рівнів фактору плану. Послідовне переміщення на кожному кроці вершини многогранника, вартість установки якої на інший рівень мінімальна, надає можливість отримати оптимальний або близький до нього за вартістю план експерименту.

Табл.4.14. Вартості змінювання значень рівнів факторів

Вартості змін.	X_1	X_2	X_3
$+1 \rightarrow -1$	3	5	7
$-1 \rightarrow +1$	5	2	5

Табл.4.15. Вихідний і оптимальний план експерименту

Вихідний план				Оптимальний план			
N	X_1	X_2	X_3	N	X_1	X_2	X_3
1	-1	-1	-1	4	-1	+1	+1
2	-1	-1	+1	2	-1	-1	+1
3	-1	+1	-1	6	+1	-1	+1
4	-1	+1	+1	8	+1	+1	+1
5	+1	-1	-1	7	+1	+1	-1
6	+1	-1	+1	5	+1	-1	-1
7	+1	+1	-1	1	-1	-1	-1
8	+1	+1	+1	3	-1	+1	-1

Оптимізація цього плану здійснюється у такій послідовності:

1. Вибирається рядок, перехід на який максимальний за вартістю.

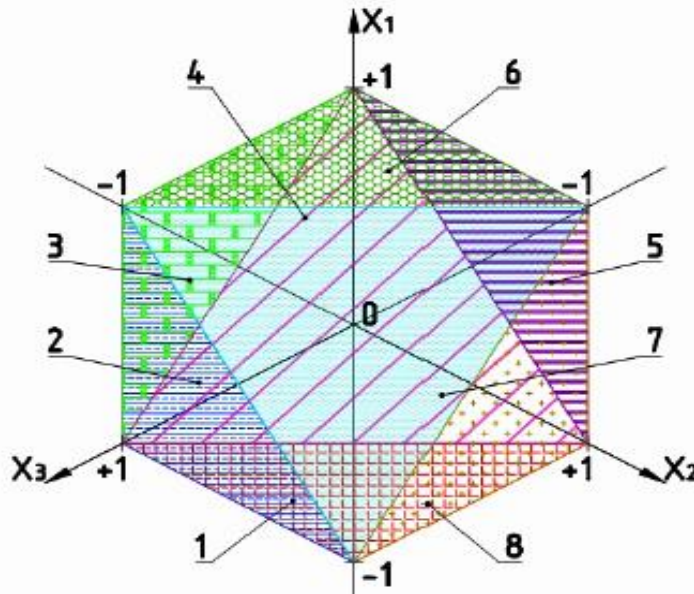


Рис. 4.5 Графічне подання реалізації методу для плану 2^3 , [21]

2. Вибір вершини, переміщення якої мінімальне за вартістю. Переміщуємо вершину $X_2 \rightarrow -1$, отримуємо новий рядок плану (трикутник переходить у позицію 2): $(-1, -1, +1)$.

3. Переміщуємо вершину X_1 , оскільки вартість її переміщення найменша поміж решти вершин. Трикутник переходить у нову позицію 3 $(+1, -1, +1)$.

4. Переміщення усіх 3-х вершин рівнозначне за вартістю, обираємо першу за номером X_1 , її переміщення призводить до появи рядку, що уже був, тому переміщуємо X_2 , отримуємо новий рядок плану, трикутник переходить у позицію 4 $(+1, +1, +1)$.

5. Переміщуємо мінімальну за вартістю переміщення вершину X_2 , отримуємо рядок, що повторюється. Переміщуємо наступну за зростанням вартості вершину. Вершини X_1 і X_3 рівнозначні за вартістю переміщення, обираємо першу за порядковим номером X_1 , отримуємо знову рядок, що повторюється. Тому переміщуємо вершину X_3 , отримуємо новий рядок плану. Трикутник переходить у позицію 5 $(+1, +1, -1)$.

6. Переміщуємо мінімальну за вартістю переміщення вершину X_2 , отримуємо новий рядок плану. Трикутник переходить у позицію 6 $(+1, -1, -1)$.

7. Переміщення вершин X_1 і X_2 рівнозначне за вартістю, обираємо першу за номером X_1 , отримуємо новий рядок плану, трикутник переходить у позицію 7 $(-1, -1, -1)$.

7. Переміщуємо мінімальну за вартістю переміщення вершину X_1 , отримуємо рядок, що повторюється. Переміщуємо наступну за зростанням вартості вершину X_2 , отримуємо новий рядок плану, трикутник переходить у позицію 8 $(-1, +1, -1)$.

Вартість реалізації вихідної матриці складає 58 у.о., оптимальна вартість реалізації, що отримана методом повного перебору (40320 варіантів) складає 25 у.о., вартість реалізації плану експерименту, що отримано викладеним методом, складає 27 у.о.

Алгоритм реалізації:

1. Об'ява змінних, що використовуються, і присвоєння їм початкових значень.

2. Зчитування із вхідного файлу даних про кількість факторів, матриці вихідного плану експерименту і вартості змінювань значень факторів.
3. Виконується оцінка вартості реалізації вихідної матриці планування.
4. Вибирається рядок, перехід на який має максимальне вартісне значення.
5. Виконується генерування матриці варіантів значень рівнів, в які можуть перейти фактори, та відповідної матриці вартостей переходів. У матриці вартостей здійснюється пошук найменшого значення. Новий рядок плану записується як попередній із перейменуванням відповідного фактору на рівень із мінімальною вартістю переходу.
6. Перевіряється значення змінної, що підсумовує кількість рядків, що повторюються, в які можна було б перейти.
7. Здійснюється перевірка наявності рівноцінних за вартістю переміщення рівнів факторів.
7. У випадку наявності у рядку рівноцінних за вартістю переміщення рівнів факторів у матрицю повторень заноситься інформація про номер рядка, значення вартості, кількості його повторень та порядковому номеру повторення, що обране.
8. Перевіряється наявність рядку, що отримано, у вихідній матриці. Якщо він є, у матрицю виключень заноситься інформація про позицію рівня фактору, в який не можна перейти із даного рядка, і здійснюється перехід на крок 5. Якщо ні, рядок, що отримано, заноситься у вихідну матрицю.
9. Якщо на кроці 6 змінна $scet$ має значення

$$scet = k(koliz - 1),$$
 де k -- кількість факторів, $koliz$ -- кількість рівнів, тобто усі рядки, в які можна перейти, уже містяться у вихідній матриці, здійснюється повернення до рядку, що містить рівноцінні за вартістю переміщення рівні факторів.
10. Вибирається наступний за порядковим номером варіант. Рядок матриці планування перезаписується із переміщенням обраного рівня фактору.
11. Перевіряється наявність отриманого рядку у вихідній матриці. За негативного результату отриманий рядок заноситься у вихідну матрицю, інакше перехід на крок 13.
13. Перевіряється, чи є переміщений на кроці 11 рівень фактору останнім із рівноцінних за вартістю переміщення. За негативного результату здійснюється перехід на крок 11 зі збільшенням порядкового номера варіанту на одиницю. Інакше здійснюється перехід на крок 10 до попереднього рядку, що містить рівноцінні за вартістю переходу рівні фактору.
14. Здійснюється запис отриманого рядку у вихідну матрицю планування.
15. Виконується перевірка досяжності потрібної кількості рядків вихідної матриці. Якщо вихідна матриці не досягла потрібної кількості рядків, перехід на крок 5.
16. Оцінюється вартість реалізації вихідної матриці планування.
17. У вихідний файл заносяться дані про вхідну та вихідну матриці планування, вартості їхньої реалізації, співвідношення цих вартостей.

4.13 Задачі.

Побудувати канонічні рівняння поверхонь

1.

$$Y = X_1^2 + 3X_2^2 + 2X_1X_2.$$

2.

$$Y = 100(X_2 - X_1)^2 + (1 - X_1)^2.$$

3.

$$Y = 1 - 2X_1 - 2X_2 - 4X_1X_2 + 10X_1^2 + 2X_2^2.$$

4.

$$Y = 3X_1^2 + 5X_2^2 + 4X_1X_2.$$

5.

$$Y = X_1^2 + 6X_1X_2 - 4X_1 - 2X_2.$$

6.

$$Y = X_1^2 + X_2^2 + X_3^2.$$

7.

$$Y = (X_1 - 1)^2 + (x_2 - 3)^2 + 4(X_3 + 5)^2.$$

7.

$$Y = 3(X_1 - 3)^2 + 2(X_2 - 2)^2 + (X_3 - 3)^2.$$

8.

$$Y = 3(X_1 - 4)^2 + 5(X_2 + 3)^2 + 7(2X_3 + 1)^2.$$

4.14 Контрольні запитання

1. Для вирішення яких задач застосовують планування експериментів.
2. За яким критерієм визначається адекватність регресійної моделі?
3. За яким критерієм визначається значимість коефіцієнтів регресійної моделі?
4. У чому полягає сутність метода найменших квадратів?
5. У чому полягає різниця між пасивним і активним експериментом?
6. У чому полягає різниця між повним факторним експериментом та дробовим?
7. У чому полягає сутність багатофакторного регулярного плану експерименту?
7. Що таке повний факторний експеримент?
8. Дайте визначення ротатбельного планування експерименту.
9. Дайте визначення центрального композиційного ротатбельного плану експерименту.
10. Які методи використовують для вирішення задачі оптимального планування експерименту?

Розділ 5 Метод групового урахування аргументів

5.1 Вступ

Під час обробки результатів експериментів важливою є проблема оцінки структури моделі об'єкту, що досліджується.

Метод групового урахування аргументів (МГУА) ґрунтується на самоорганізації моделей і спрямований на визначення структури моделей оптимальної складності.

У МГУА входять параметричні алгоритми, алгоритми кластеризації, комплектування аналогів тощо. За допомогою перебору різних рішень в індуктивному підході прагнуть мінімізувати роль впливу аналітика на результати моделювання.

Нехай об'єкт подано у вигляді сукупності станів $S_1, S_2, \dots, S_k$, за якими можна судити про його поведінку всередині системи або за її межами.

Кожний стан характеризується набором параметрів $x_1, x_2, \dots, x_n, x_{n+1}$:

$$S_i = S_i(x_1, x_2, \dots, x_n, x_{n+1}).$$

У разі статичної моделі стан системи визначається регресійною моделлю. У разі динамічної моделі стан системи визначається, наприклад, диференціальним рівнянням або системою рівнянь.

Поміж параметрів, що розглядаються, відокремлюють кілька найзначиміших, які визначають мету дослідження. Саме такий сенс вкладають у термін результативна ознака і позначають його Y_0 . Нехай таким параметром є x_{n+1} , тобто $Y_0 = x_{n+1}$.

Ставиться задача визначення математичної залежності результативної ознаки Y_0 із набором параметрів, що управляються, $(x_1, x_2, \dots, x_n)$, які вважаються незалежними.

Побудова і аналіз ефективного рівняння $Y = \varphi(x_0)$ є основний момент у теорії статистичних методів, оскільки саме результативна ознака використовується у подальшому для управління об'єктом, виведення системи на оптимальний рівень, прогнозування за межами системи і реалізації інших прикладань.

Для створення вхідних даних для МГУА використовують різниці рівняння.

За малим обсягом вибірки коли при побудові регресійних моделей немає можливості використовувати статистичні гіпотези про густину розподілу, використовують індуктивний підхід.

Згідно цьому методу послідовно породжуються моделі зростаючої складності доти, доки не відшукається мінімум деякого критерію якості моделі. Цей критерій зветься зовнішнім критерієм, оскільки під час налаштування моделей і оцінці якості моделей використовують різні дані. Досягнення глобального мінімуму зовнішнього критерію при породженні моделей означає, що модель, яка доставляє мінімум, є шукана.

5.2 Постановка задачі

Для побудови моделі оптимальної складності розв'язується така задача: за даними матриці спостережень вхідних змінних X розміру $n \times m$, де n -- число спостережень m -- число вхідних змінних, і вектору спостережень вихідної змінної y розмірності n необхідно оцінити залежність (структуру моделі та її параметрів)

$$y = X\theta^0 + \xi = y^0 + \xi, \quad (1)$$

де y^o -- точний або істинний вихід об'єкту (сигнал); θ^o -- невідомий істинний вектор параметрів, структуру і значення якого треба оцінити; ξ -- вектор випадкових завад (шум); X -- матриця входних змінних.

При цьому треба сформувати деяку множину $\mathcal{J}$ моделей різної структури у вигляді $y_f = f(X\hat{\theta}_f)$ і знайти оптимальну модель f^* із мінімальним значенням заданого критерію $CR(\cdot)$:

$$f^* = \operatorname{argmin}_{f \in \mathcal{J}} CR(y, f(X, \hat{\theta}_f)), \quad (2)$$

де $\hat{\theta}_f$ -- вектор оцінок параметрів для кожної $f \in \mathcal{J}$, що визначається як розв'язок неперервної оптимізаційної задачі

$$\hat{\theta}_f = \operatorname{arg} \min_{\theta_f \in R^{s_f}} Q(y, X, \theta_f), \quad (3)$$

де s_f -- складність моделі f ; $Q(\cdot) \neq CR(\cdot)$ -- критерій якості розв'язку задачі параметричної ідентифікації кожної частинної моделі, що генерується у процесі розв'язання задачі структурної ідентифікації (1).

Якість моделі можна визначати за значеннями критерію регулярності $AR(f)$, який будується на основі розбиття вибірки X на дві підвибірки X_A і X_B . На *навчальній підвибірці* X_A оцінюються параметри моделі, а на *перевірочній підвибірці* X_B обчислюють похибку

$$AR(f) = \|y_B - f(X_B, \hat{\theta}_{Af})\|^2, \quad (4)$$

де $\hat{\theta}_{Af}$ -- вектор параметрів моделі f , що визначається на підвибірці X_A .

Якість моделі f^* вважається оптимальною, коли значення критерію $AR(f)$ буде найменшим, тобто

$$f^* = \operatorname{argmin}_{s=1, m} AR(f). \quad (5)$$

За використання комбінаторного алгоритму виконується перебір усіх можливих моделей із вибором найкращої за критерієм селекції. Загальна кількість P_m усіх можливих моделей, які містять від 1 до m аргументів, дорівнює

$$P_m = \sum_{j=1}^m C_m^j = 2^m - 1. \quad (6)$$

У структурі комбінаторного алгоритму можна відокремити такі основні блоки:

- 1) перетворення початкових даних відповідно до обраної системи опорних (базисних) функцій, в якій шукається модель (базис може бути поліноміальним, тригонометричним, у формі різницевих рівнянь тощо);
- 2) генерування і перебір моделей різної структури у обраному базисі;
- 3) обчислення значень деякого критерію селекції, який має властивості зовнішнього доповнення, і послідовний вибір частинних моделей, кращих за цим критерієм.

Найбільший істотний вплив на швидкодію алгоритму має другий блок *генерування (перебір) частинних моделей*. Основні операції, що здійснюються у цьому блоці, такі:

формування структури чергової частинної моделі та відповідної системи

нормальних рівнянь, а також її розв'язання (оцінка коефіцієнтів моделі).

Структура частинних моделей формалізується за допомогою структурного вектора $d = (d_1, d_2, \dots, d_m)$: якщо елемент d_i набуває значення 1, відповідний i -й аргумент включається у модель. При здійсненні повного перебору моделей найчастіше використовується схема зміни вектору d за принципом двійкового лічильника. При її використанні у частинну модель включаються аргументи, що відповідають стану двійкового регістру, що встановлений після додавання одиниці в останній розряд (число розрядів m дорівнює числу аргументів).

Ця схема надає можливість організувати рекурентну процедуру перебору, за якої досягається найкраще узгодження схеми перебору моделей і методу обчислення оцінок коефіцієнтів. Вона також надає можливість досить просто виключити із розгляду моделі із лінійно залежними членами.

Простим за ідеєю, але складнішим для реалізації, є *лічильник з послідовним ускладненням моделей, послідовний лічильник*, який доцільно застосовувати в алгоритмах повного перебору з обмеженням складності моделей. При цьому спершу розглядаються усі можливі варіанти розміщення у векторі d однієї одиниці ($C_m^1 = m$ варіантів), потім двох і так далі, до m одиниць ($C_m^m = 1$ варіант).

5.3 Багаторядний МГУА

Нехай в якості початкових даних вибрано матрицю $A = (X_1, X_2, \dots, X_n, Y)$, де X_i , $i = \overline{1, n}$, Y -- вектори-стовпці розмірності m , X_i -- вхідні фактори; Y -- результуюча характеристика. Задача полягає в ідентифікації залежності

$$Y = F(X_1, X_2, \dots, X_n) \quad (7)$$

поліномом Колмогорова--Габора

$$Y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i < j}^n a_{ij} x_i x_j + \sum_{i < j < k}^n a_{ijk} x_i x_j x_k + \dots \quad (8)$$

Особливістю МГУА є можливість його застосування у випадку малої кількості точок експерименту.

На першому етапі реалізації МГУА обирається опорна функція виду

$$1. y = a_0 + a_1 x_i x_j;$$

$$2. y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j;$$

$$3. y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j.$$

Для першої функції потрібно дані бодай трьох експериментів, для другої -- 5, для 4-ї -- 7. Позначимо $y_k = f(x_i, x_j)$, де f -- одна із казаних залежностей або подібна до неї.

На наступному кроці за допомогою МНК визначають коефіцієнти рівнянь

$$y_1 = f(x_1, x_2); y_2 = f(x_1, x_3), \dots, y_{n-1} = f(x_1, x_n),$$

$$y_n = f(x_2, x_3), \dots, y_p = f(x_{n-1}, x_n), \quad (9)$$

де $p = X_n^2$.

Після ідентифікації усіх залежностей y_i , $i = \overline{1, p}$, за зовнішнім критерієм обирають

кращі. Залежності, що залишилися, перенумеровують: $y_1, y_2, \dots, y_s$, s -- кількість відібраних залежностей. Перший крок селекції закінчено.

На наступному кроці за допомогою МНК визначають коефіцієнти таких залежностей:

$$z_1 = f(y_1, y_2); z_2 = f(y_1, y_3), \dots, z_r = f(y_{s-1}, y_s), r = C_s^2. \quad (10)$$

Подальша процедура аналогічна наведеній. Якщо значення зовнішнього критерію покращується, селекція триває, інакше отримано модель оптимальної складності.

Залежно від типу задачі А.Г. Івахненко запропонував розглядати такі критерії: регулярності, незміщеності та балансу змінних.

5.3.1 Критерії регулярності

Відомі два критерії регулярності:

-- мінімум середньоквадратичної похибки на нових точках окремої контрольної послідовності;

-- максимум коефіцієнту кореляції на тих самих точках.

Розглянемо процедуру їх застосування. Початкові дані містяться у відокремленій частині табл. 5.1.

Табл. 5.1. Вхідна інформація для ідентифікації за допомогою МГУА

x_1	x_2	$\dots$	x_n	y	y_1	y_2	$\dots$	y_p
x_{11}	x_{12}	$\dots$	x_{1n}	y_1	y_{11}	y_{12}	$\dots$	y_{1p}
x_{21}	x_{22}	$\dots$	x_{2n}	y_2	y_{21}	y_{22}	$\dots$	y_{2p}
$\cdot$	$\cdot$	$\dots$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\dots$	$\cdot$
x_{m1}	x_{m2}	$\dots$	x_{mn}	y_m	y_{m1}	y_{m2}	$\dots$	y_{mp}

Розділимо її рядки на дві частини (приблизно 60% на 40%), тоді $m = l + k$, l -- кількість точок експерименту у першій (навчальній вибірці, k -- у другій (контрольній). Значення l має бути більше числа складових в опорній функції $f(x)$.

Із використанням початкової вибірки визначаємо коефіцієнти залежностей (9). Потім розраховуємо значення критерію регулярності на точках контрольної вибірки, після чого упорядковуємо y_i за збільшенням критерію і вибираємо із визначену кількість із найменшим його значенням. Після перенумерації вони складуть множину функцій наступного ряду селекції. Умови закінчення ітерацій можуть бути такими:

- середнє значення похибки для наступного ряду селекції є більшим за найбільше (середнє) значення похибки для попереднього ряду;
- мінімальне значення похибки наступного ряду селекції є більшим за мінімальне значення похибки попереднього ряду;
- максимальне значення похибки наступного ряду є більшим за максимальне значення похибки попереднього ряду;
- модуль відхилення похибок наступного і попереднього рядів менше за деяке число.

Критерій регулярності -- мінімізація СКП (середньоквадратичної похибки) на точках контрольної послідовності:

$$\varepsilon = \frac{\sum_{i=1}^{N_k} (y_i - \hat{y}_i)}{\sum_{i=1}^{N_k} y_i^2} \rightarrow \min, \quad (11)$$

де N_k -- кількість точок у контрольній послідовності; $\hat{y}_i$ -- значення шуканої залежності в i -й точці контрольної послідовності.

Другий варіант критерію регулярності полягає у максимізації коефіцієнту кореляції

$$k = \frac{\sum_{i=1}^{N_k} (y_i \hat{y}_i)}{\sum_{i=1}^{N_k} y_i^2} \rightarrow 1. \quad (12)$$

Для отримання результатів, що співставимі для різних моделей, величини (11), (12) нормують.

Перевагою критерію регулярності є плавність змінювання його значення за збільшення складності моделі. Недоліком його використання є низька точність при вирішенні екстраполяційних задач. Тому цей критерій слушно застосовувати для ідентифікації і коротко термінового прогнозу.

5.3.2 Критерій незміщеності

Відомі три види критерію незміщеності, що базуються на аналізі рішень (КН1), аналізі коефіцієнтів (КН2) і "критерій відносної незміщеності".

КН1. Тут необхідно ранжувати усі точки експериментів за збільшенням чи зменшення значення дисперсії. Після ранжування точки експериментів нумерують і ділять на дві послідовності:

- до 1-ї відносять точки із парними номерами -- N_1 ;
- до 2-ї відносять точки із непарними номерами -- N_2 ; $N_1 + N_2 = n$.

На першому ряді селекції перша послідовність є навчачою, друга -- контрольна. Рівняння регресії, що отримані на навчачій послідовності N_1 позначимо $y_k^1 = f(x, x_j)$. Далі першу послідовність вважають контрольною, другу -- навчачою. На навчачій послідовності отримують рівняння регресії $y_k^2 = f(x, x_j)$. Кількість рівнянь y_k^1 та y_k^2 має збігатися. Для кожного k розраховують СКВ (середньоквадратичне відхилення)

$$n_k = \frac{1}{m} \sqrt{\sum_{i=1}^m (y_{k_i} - y_{k_i}^2)^2}, \quad (13)$$

k_i -- номер i -го рівняння. Вибирають p рівнянь, які відповідають меншій оцінці n_k (y_k^1 чи y_k^2). Середнє значення критерію незміщеності на першому ряді селекції обчислюють за формулою:

$$N_1 = \frac{1}{p} \sum_{i=1}^p n_i. \quad (14)$$

На другому і наступних рядах селекції процедура та сама. Селекція триває доти, доки середнє значення критерію незміщеності зменшується.

КН2. Точки експериментів ранжують за величиною дисперсії і ділять на навчаючу і контрольну послідовності навпіл. Точки з більшим значенням дисперсії опиняються у навчаючій послідовності, з меншим -- у контрольній. Особливість критерію полягає у тому, що на кожному ряді селекції ранжування і поділ точок експерименту здійснюється заново. Крім того, зменшується свобода вибору відповідно формулі $F = \alpha t_\beta$, де F -- число змінних, що пропущено на цьому ряду селекції, S -- номер ряду, t -- кількість вхідних змінних; $\alpha = 1 \div 5$, $\beta = 0,1 \div 0,2$. Формула і процедура, що пов'язана із нею, надає можливість найшвидшого вирішення задач великої розмірності.

Значення критерію незміщеності оцінок коефіцієнтів моделі розраховують за формулою

$$n_k = \frac{\sum_{i=1}^p \chi(a_i = b_i) \cdot a_i^2}{\sum_{i=1}^p a_i^2 + \sum_{i=1}^p b_i^2}, \quad (15)$$

де p -- загальна кількість коефіцієнтів, a_i -- коефіцієнти, що отримано до змінювання послідовностей, b_i -- після. У наступний ряд пропускають F моделей, які мають більше значення n_k . Селекція ривкає доти, доки n_k збільшується.

Критерій незміщеності треба застосовувати або разом із критерієм регулярності, або в алгоритмах із повним перебором моделей.

Критерій відносної незміщеності. У цьому разі використовують тільки лінійні частинні описи (моделі), але для запобігання втрати точності простір початкових аргументів містить і коваріації (наприклад, $x_2, x_2, x_1 x_2 \rightarrow x_1, x_2, x_3$). Оскільки частинний опис на другому ряду є:

$$z = a_0 + a_1 y_i + a_2 y_j,$$

ортогоналізований частинний опис:

$$z = y_i + A \hat{y}_j, \quad (16)$$

де $\hat{y}_j$ -- вектор, ортогональний по відношенню до x_i .

Якщо у частинних описах значення змінних центровані і нормовані за середнім значенням, в ортогоналізованих частинних описах вільний член $a_0 = 0$, а інші коефіцієнти мають такі значення:

$$a_1, a_2 = A = \frac{\sum_{i=1}^N (y_i y_j)}{\sum_{i=1}^N y_i^2}. \quad (17)$$

Ортогоналізація є перетворенням $\hat{y}_j$ відносно базового y_i , у результаті якого отримуємо:

$$\sum_{i=1}^N y_i \hat{y}_j = 0.$$

Для цього досить, щоб $\hat{y}_j = y_j - A y_i$.

Отже, алгоритм МГУА із використанням критерію відносної незміщеності має такі базові кроки:

1. Початкові дані ділимо на дві частини.
 2. На першій послідовності визначаємо значення коефіцієнтів A^* у рівнянні регресії, на другій -- коефіцієнти A^{**} .
- Кращим вважаємо ті описи, в яких

$$n_{zm} = \frac{A_i^* - A_i^{**}}{A_i^*} = \min n_{zm}.$$

Ряди описів будуть такими:

$$y = a_1 x_i; \quad z = y_i + A \hat{y}_j; \quad t = z_i + A \hat{z}_j, \dots$$

5.3.3 Критерій балансу змінних

Критерій балансу змінних є найефективніший критерій за середньо- і довгостроковому прогнозуванні. Його визначення може бути емпіричним та штучним.

За емпіричного визначення критерію балансу змінних із перебирання виключаємо моделі, що дають явним чином невірний прогноз (тобто значення Y , що прогнозується, не належить області, якій має належати за своїм визначенням). Крім того, на Y може бути накладено безліч інших умов. Наприклад, у залежності $y_k = f(x_i x_j)$ $y_k^l < y_k^{l+1}$, $l = \overline{1, n-1}$, n -- кількість точок, які ураховуються при побудові y_k , або $y_k^l / y_i^l \in [1; 3]$, $l = \overline{1, n-1}$.

Штучні умови балансу не є наслідком принципу фізичної реалізованості моделі і визначаються дослідником. Зазвичай, це -- функції, що є комбінації сум і різниць значень вхідних факторів, при цьому у систему вхідних факторів додаються незалежні змінні, що зазвичай на мають фізичного сенсу. Наприклад, у задачу прогнозування трьох змінних X_1, X_2, X_3 запроваджуються нові змінні

$$S_1 = X_1 + X_2 + X_3; S_2 = -X_1 + X_2 + X_3; S_3 = -X_1 - X_2 + X_3;$$

$$S_4 = X_1 - X_2 - X_3; S_5 = -X_1 - X_2 + X_3; S_6 = -X_1 - X_2 - X_3. \quad (18)$$

Ці змінні разом із початковими змінними X_1, X_2, X_3 утворюють вектор вхідних факторів і прогнозуються разом з іншими змінними.

Початкові дані для МГУА

t	X_1	X_2	X_3	S_1	S_2	$\dots$	S_6
t_1	x_{11}	x_{12}	x_{13}	s_{11}	s_{12}	$\dots$	s_{16}
t_2	x_{21}	x_{22}	x_{23}	s_{21}	s_{22}	$\dots$	s_{26}
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\dots$	$\cdot$
t_n	x_{n1}	x_{n2}	x_{n3}	s_{n1}	s_{n2}	$\dots$	s_{n6}

Припустимо, що у результаті поступового ускладнення моделей і застосування МНК отримано рівняння трендів

$$X_j^k(t) = a_{0j}^k + \sum_{i=1}^k a_{ji}^k t^i, \quad j = \overline{1, 3}, k = \overline{1, m}. \quad (19)$$

Найкращою вважається така система трендів, що має мінімальне розсіювання в інтервалі часу спостереження

$$\Phi = \sum_{l=1}^6 E_l^2 = \sum_{l=1}^6 \sum_{t=t_1}^{t_n} (S_l(t) - \hat{S}_l(t))^2$$

$$= \sum_{l=1}^6 \sum_{t=t_1}^{t_n} \sum_{k_2, k_3=1}^m (S_l(t) - \hat{S}_l(X_1^{k_1}(t), X_2^{k_2}(t), X_3^{k_3}(t)))^2 \rightarrow \min, \quad (20)$$

де $\hat{S}_l(*)$ -- залежності типу (18).

Тренди -- рішення цієї задачі -- є оптимальні залежності.

Вигляд задачі вказує на суттєву трудомісткість процесу її вирішення. Тому треба використовувати емпіричні процедури зменшення кількості комбінацій перебирання, у тому числі враховувати принцип можливості фізичної реалізованості. Число m обирається дослідником залежно від потужностей комп'ютера.

5.3.4 Алгоритм розділення початкової вибірки даних

Найрозповсюдженіший є підхід, за яким у навчаючу вибірку вибирають точки із більшим значенням дисперсії, а у контрольну -- з меншими. Це пояснюється тим, що область навчання має бути широкою, а контрольні точки, зазвичай, міститися всередині цієї області.

Алгоритм розділення:

1. Визначити відсоткове співвідношення між кількістю елементів у навчаючій та контрольній послідовностях.

2. Для кожного стовпця X_i , $i = \overline{1, n}$, розрахувати середнє значення його елементів

$$x_{i_{\text{ср}}} = \frac{1}{m} \sum_{j=1}^m x_{ji}$$

та отримати середнє значення множини $X_{n_{\text{ср}}} = (x_{1_{\text{ср}}}, x_{2_{\text{ср}}}, \dots, x_{n_{\text{ср}}})$.

3. Знайти вибіркві дисперсії для кожного рядку таблиці з формулою

$$D_i = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - x_{j_{\text{ср}}})^2, \quad i = \overline{1, m}.$$

4. Для упорядкування елементів таблиці переставити рядки так, щоб першою був рядок із найбільшим значенням дисперсії.

5. Відповідно до результату кроку 1 розділити дані таблиці на навчаючу та контрольну послідовності.

5.4 Опис індуктивного алгоритму

Індуктивний підхід на основі МГУА передбачає перебір багатьох можливих варіантів розв'язання задачі моделювання і вибір найкращого варіанту моделі. Щоб побудувати найадекватнішу модель, у вибірку додають якомога більше аргументів. Але за великої їх кількості перебір усіх варіантів займає багато часу і часто стає неможливим. Тому виникає проблема прискорення комбінаторного алгоритму МГУА за рахунок відбору інформативних аргументів, які мають найбільший рівень впливу на вихідну змінну, і видалення неінформативних із вибірок.

Неінформативні аргументи мають малі коефіцієнти в моделях, внаслідок чого вони

мають слабкий вплив на значення вихідної змінної. Для цього запропоновано використання так званого *вагового критерію WC* оцінки ступеню інформативності аргументу на множині $\mathcal{J}_F$:

$$WC(x_j, \theta_{jf}, CR(f)) = \frac{\|x_j\|^2}{nF} \sum_{k=1}^F \frac{\theta_{jf}^2}{CR(f_k)}. \quad (21)$$

Тут $\|x_j\|$ -- норма вектору вимірювань аргументу у вибірці W , F -- свобода вибору; θ_{jf_k} -- коефіцієнт при аргументі x_j у моделі f_k .

Критерій WC враховує вагу у кожній моделі та якість $CR(f)$ моделі, в яку він входить.

Усі розглянуті способи генерування структур моделей перебірних алгоритмів можна подати у вигляді схеми рис. 5.1.



Рис. 5.1 Класифікаційна схема перебірних алгоритмів

Індуктивний алгоритм визначення моделі оптимальної структури складається із таких основних кроків, [35]:

Крок 1. Нехай задана вибірка

$$D = \{(x_n, y_n)\}_{n=1}^N, \quad x \in R^P.$$

Вибірка ділиться на тестову (L) та навчальну (C). Ці множини задовольняють умовам $L \cup C = W$, $L \cap C = 0$. Матриця X_L складається із тих векторів-рядків x_n , для

яких індекс $n \in L$. Вектор Y_L складається із тих елементів y_n , для яких індекс $n \in L$. Зазвичай вибірка, що навчає, складає 70% із всієї вибірки.

Внаслідок розбиття отримуємо

$$X_w = \begin{pmatrix} X_L \\ X_C \end{pmatrix}, \quad Y_w = \begin{pmatrix} Y_L \\ Y_C \end{pmatrix}, \quad |L| + |C| = W.$$

Крок 2. Призначається базова модель. Під базовою моделлю мається на увазі вид відношення між залежною змінною y та вільними змінними x із базисним набором опорних функцій, що використовуються у подальшому для формування моделей різної складності.

Крок 3. Виходячи із поставленої задачі вибирається зовнішній критерій, що описує якість моделі.

Крок 4. Породжуються моделі-претенденти на основі отримання частинних описів із базової моделі (схрещування) і відбору кращих із них (селекція).

Крок 5. Налаштовуються параметри моделей. Для цього використовують внутрішній критерій -- критерій, що обчислюється із використанням вибірки, що навчає. Кожному елементу вектора x_n -- елементу вибірки D ставиться у відповідність вектор a_n . Будується матриця A_w -- набір векторів-стовпців a_n . Матриця A_w розбивається на підматриці A_L і A_C .

Вважається, що

$$\|y_C - y_{C_0}\|^2 = (y_C - y_{C_0})^T (y_C - y_{C_0}), \quad y_{C_0} = A_C w_{C_0}$$

доставляє значення вектору параметрів w_{C_0} , що обчислюється за МНК:

$$w_{C_0} = (A_C^T A_C)^{-1} A_C^T y_C.$$

Крок 6. Обчислюється якість породжених моделей. При цьому використовується тестова вибірка і назначений зовнішній критерій.

Крок 7. Модель, що доставляє мінімум зовнішньому критерію, вважається оптимальною

Якщо значення зовнішнього критерію не досягає свого мінімуму за збільшення складності моделі або значення функції якості незадовільне, обирається краща модель із заданої складності. Під складністю моделі мається на увазі кількість параметрів, що налагоджуються. Існують такі причини, за яких глобальний мінімум може не існувати:

- 1) дані занадто зашумлені;
- 2) поміж даних нема необхідних для відшукування моделі змінних;
- 3) при аналізі часових рядів існує значна часова затримка причинно- наслідкового зв'язку, що відшуковується.

5.4.1 Зовнішні критерії

Критерій вибору моделі можна назвати зовнішнім, коли він отриманий за допомогою додаткової інформації, що не міститься у даних, які використовувалися при обчисленні параметрів моделей. Наприклад, така інформація міститься у додатковій, тестовій вибірці.

Внутрішній критерій використовується для налагодження параметрів моделі, а зовнішній -- для вибору моделі оптимальної структури. Можливий вибір моделі за кількома критеріями.

Найрозповсюдженіші критерії: Критерій регулярності і критерій мінімального

зміщення.

Критерій регулярності $\Delta^2(C) = \Delta^2(C \setminus L)$ містить середньоквадратичну похибку на підвибірці C , що навчає, яка отримана при параметрах моделі, що налагоджені на тестовій підвибірці L :

$$\Delta^2(C) = \|y_C - A_C w_{L_0}\|^2.$$

Критерій мінімального зсуву (критерій непротиріччя) моделі: модель, що має на вибірці, що навчає, одну нев'язку, а на тестовій -- іншу, називається моделлю із протиріччям. Цей критерій включає різницю між залежними змінними моделі, що обчислюються на двох різних вибірках L і C . Критерій не включає похибку моделі в явній формі. Він потребує, щоб оцінки коефіцієнтів в оптимальній моделі, що обчислені на множинах L і C , відрізнялися мінімально.

Критерій непротиріччя як критерій мінімуму зсуву має вигляд:

$$\eta_\alpha^2 = \|w_{L_0} - w_{C_0}\|^2.$$

5.5 Комбінаторний алгоритм МГУА

У комбінаторних алгоритмах виконується перебір усіх можливих частинних моделей із заданого поліноміального базису із вибором кращої із цих моделей за заданим зовнішнім критерієм. Під час перебору складність частинних моделей нарощується від 1 до максимального числа n (число аргументів базисного набору функції). Отже, загальна схема комбінаторного алгоритму містить такі операції: за методом МНК визначаються коефіцієнти усіх частинних моделей при складності $s = 1, s = 2, \dots, s = n$; для кожної із них обчислюється значення зовнішнього індивідуального і комбінаторного критерію селекції, єдина оптимальна модель вибирається за мінімальним критерієм.

Структура комбінаторного алгоритму, [16]. У структурі алогритму відокремлюються три основних блоки:

1. Перетворення вхідних даних відповідно вибраній системі опорних (базисних) функцій, в якій відшукується модель.
2. Генерування (перебір) повної або неповної множини частинних моделей, що ускладнюються, в обраному базисі.
3. Обчислення значення деякого критерію селекції, що має властивості зовнішнього доповнення, та послідовного відбору частинних моделей, кращих за цим критерієм.

Формування базису.

За заданими вимірюваннями деяких вхідних змінних $z_1, z_2, \dots, z_m$ об'єкту, що моделюється, і максимального степеню поліному, число доданків n у повному поліномі степеню $k_{\max}$ від m змінних визначається однозначно:

$$n = \prod_{i=1}^m (k_{\max} + i) / i.$$

Повний поліном моделі:

$$y = \sum_{i=1}^n a_i x_i,$$

де кожний узагальнений лінійний аргумент x_i є нелінійна функція вхідних змінних z_j :

$$x_i = \prod_{j=1}^m z_j^{k_{ij}}.$$

Для формування степенів аргументів зручно організувати процедуру отримання послідовності m -розрядних чисел $k_i = (k_{i1}, k_{i2}, \dots, k_{im})$ із основами $1, 2, \dots, k_{\max} + 1$ і відбирати ті з них, які задовольняють указаному обмеженню. Наприклад, за $m = 2$ (два параметри) і $k_{\max} = 2$ отримаємо послідовність $n = 6$ чисел k_i , $i = \overline{1, 6}$: 00, 10, 01, 20, 11, 02, що відповідає повному поліному

$$y = a_1 + a_2 z_1 + a_3 z_2 + a_4 z_1^2 + a_5 z_1 z_2 + a_6 z_2^2.$$

Отже, члени x_i повного поліному є базисний набір опорних функцій для комбінаторного алгоритму. Описана процедура надає можливість сформувати матрицю вимірювань узагальнених аргументів $X[N \times n]$, де N -- число точок вимірювань.

Матрицю X слід розділити на матрицю C , що навчає, довжиною N_C і тестову L довжиною N_L , $N_C + N_L = N$.

Генерування (перебір) частинних моделей

Основними операціями, що виконуються у цьому блоці, є:

- формування структури чергової частинної моделі;
- формування відповідної нормальної системи рівнянь;
- розв'язання отриманої системи (оцінка коефіцієнтів моделі).

Формування структур частинних моделей формалізується за допомогою двійкового структурного вектору $d = (d_1, d_2, \dots, d_n)$: якщо елемент d_i цього вектору набуває значення 1, відповідний i -й аргумент включається у частинну модель, якщо 0 -- ні, $i = \overline{1, n}$.

Змінювання станів вектору d можна організувати кількома способами, але найпростіший є такий: отримувати усі можливі варіанти розміщення у векторі d спочатку однієї одиниці (усього $C_n^1 = n$ варіантів), потім двох ($C_n^2 = n(n+1)/2$) варіантів тощо аж до n одиниць (один варіант). Отже, на k -й селекції число моделей, що генеруються, дорівнюватиме C_n^k .

Відповідно до наведеної схеми алгоритм функціонує так. Спочатку визначаються усі моделі за $s = 1$, тобто ті, що складаються лише із одного аргументу:

$$q_1 = a_1, \quad q_2 = a_2 x_1, \quad q_3 = a_3 x_2, \dots, q_n = a_{n+1} x_n.$$

Далі розглядаються усі можливі моделі при $s = 2$, що складаються із двох аргументів:

$$q_1 = a_1 + a_2 x_1, \quad q_2 = a_1 + a_3 x_2, \dots$$

Аналогічно будуються частинні моделі при $s = 3$, $s = 4$ тощо до $C_n^n = 1$ при $s = n$, тобто до повного поліному.

Загальна кількість варіантів складе $p_n = 2^n - 1$ різних структур, тобто повний перебір. Але програмна реалізація цього способу є далеко не швидкодіюча, і такий перебір слушно застосовувати тільки в алгоритмах неповного перебору.

За великого n доводиться застосовувати усічення перебору. Наприклад, відомий спосіб усічення трикутника перебору: розглядаються усі моделі із складністю від $s = 1$ до $s_{\max} = m$.

Найкомпактнішою та універсальною є схема змінювання двійкового вектору за принципом роботи двійкового рахівника, в останній розряд якого додається одиниця. При

цьому, що суттєво, зберігається взаємно однозначна відповідність між порядковим номером чергової частинної моделі і станом структурного вектору. Цей спосіб алгоритмічно простий і зручний саме при повному переборі. Він надає можливість організувати рекурентну процедуру перебору.

Для формування нормальної системи рівнянь, що відповідає черговому структурному вектору, можна діяти формально: із стовпців повної матриці X , із одиничними елементами d , складається частинна матриця X_1 , а потім обчислюються елементи нормальної матриці $X_1^T X_1$, $X_1^T Y$. Але за повного перебору цей формальний алгоритм є найгірший, оскільки призводить до багатократного обчислення одних і тих самих скалярних добутків. Наприклад, для структур 010,011,110 три рази обчислюється одна й та сама величина

$$x_2^T x_2 = \sum_{j=1}^N x_{2j}^2.$$

Тому у комбінаторному алгоритмі досить один раз обчислити матриці нормальної системи

$$X^T X = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 \\ x_2^T x_1 & x_2^T x_2 \end{bmatrix}$$

$$x_1^T x_n x_2^T x_1 x_2^T x_2 x_2^T x_n \cdots x_n^T x_1 x_n^T x_2 x_n^T x_n.$$

Для отримання будь-якої частинної нормальної системи досить взяти елементи матриці $X^T X$, що містяться на перетині рядків і стовпців, що вказані одиницями вектору d , а також відповідні елементи вектора $X^T Y$. Для розв'язання кожної нормальної системи $x_i^T X_i a_i = x_i^T y$, тобто для обчислення оцінок коефіцієнтів частинної моделі, можна застосувати будь-які процедури розв'язання систем алгебраїчних рівнянь.

Відбір за зовнішніми критеріями

Селекція (відбір) моделей, кращих за заданим критерієм I , зазвичай виконується у процесі перебору (а не у кінці). Для цього запам'ятовуються значення критерію для заданого числа F перших моделей, а потім величина критерію кожної наступної моделі I_i порівнюється із найгіршим $I_{\max}$ із F значень. Якщо $I_i < I_{\max}$, нова модель запам'ятовується замість найгіршої (запам'ятовуються структура, оцінки коефіцієнтів і критерій): якщо ж $I_i \geq I_{\max}$, ця модель пропускається. Після закінчення перебору F моделей, що залишилися, є кращі із усіх p_n у сенсі заданого критерію. Вочевидь, за такого підходу блоки генерації і відбору алгоритмічно об'єднуються.

В алгоритмах МГУА зазвичай виконується ще один етап обчислень -- оцінка якості відібраних кращих моделей.

Розглянемо алгоритм МГУА для оцінки структури моделі динамічного об'єкту першого порядку.

Для рівняння першого порядку різницевого рівняння має вигляд

$$y_n = a_0 x_n + a_1 x_{n-1} - b y_{n-1}.$$

Для передаточної функції $W = 2/(0,01p + 1)$ різницевого рівняння має вигляд

$$y_n = 2x_{n-1} - 0,33y_{n-1}.$$

Обчислено середньоквадратичну похибку I , виконано тестування програми, що реалізує метод МГУА.

У методі МГУА розглядаються усі можливі комбінації поліному

$$a_0 + a_1 x_n + a_2 x_{n-1} + a_3 y_{n-1} + a_4 x_n x_{n-1} + a_5 x_n y_{n-1} + a_6 x_{n-1} y_{n-1} + a_7 x_n^2 + a_8 x_{n-1}^2 + \dots$$

Для рівняння другого порядку маємо

$$y_n = a_0 x_n + a_1 x_{n-1} + a_2 x_{n-2} - b_0 y_{n-1} - b_1 y_{n-2}.$$

Для передатної функції

$$W = 3/(0,05p^2 + 0,35p + 1)$$

різницеве рівняння має вигляд

$$y_n = 0,29x_{n-1} + 0,29x_{n-2} - 1,93y_{n-1} + 0,93y_{n-2}.$$

Отримано моделі:

$$1. y_n = 0,293x_{n-1} + 0,2863x_{n-2} - 1,9305y_{n-1} + 0,9324y_{n-2};$$

$$2. y_n = 0,293x_{n-1} + 0,286x_{n-2} - 1,93y_{n-1} + 0,932y_{n-2} - 2,59e_{-13}x_{n-1}y_{n-2};$$

5.6 Контрольні запитання

1. До якого класу методів належить МГУА?
2. Які принципи закладено в основу реалізації МГУА?
3. Дайте характеристику принципу повноти в МГУА.
4. Які критерії характеризують закінчення ітераційного процесу в МГУА?
5. Яка роль завчаючої та контрольної послідовності в алгоритмах МГУА?
6. Наведіть приклади емпіричного і штучного формування критерію балансу змінних.
 7. Для вирішення яких задач слухно застосовувати критерій балансу змінних?
 7. Наведіть алгоритм поділу генеральної сукупності на навчальну і контрольну послідовності за величиною дисперсії.
 8. Яким чином в алгоритмах МГУА реалізовано елементи самоорганізації?

Розділ 6 Динамічне програмування

На відміну від принципу максимуму, який формулюється таким чином, що виявляється орієнтованим, перш за все, на визначення оптимального управління у вигляді оптимальної програми, динамічне програмування надає можливість визначати оптимальне управління тільки у формі синтезуючої функції.

Динамічне програмування добре обґрунтовано для дискретних систем і процесів. Це пов'язано із тим, що при виводі функціонального рівняння Беллмана треба робити припущення, безпосередня перевірка якого за рівняннями руху та функціоналу якості неможлива. Це можна зробити тільки після розв'язання рівняння Беллмана. Крім того, функціональне рівняння Беллмана для неперервних процесів являє собою диференціальне рівняння із частинними похідними.

Якщо мати на увазі не тільки задачі оптимального управління, динамічному програмуванню властива універсальність й його можна застосовувати для вирішення широкого кола задач оптимізації. Будемо розглядати основний зміст динамічного програмування як методу оптимізації.

Розглянемо деяку систему S (економічну, соціальну, фізичну тощо), стани якої змінюються у дискретні моменти часу $t_0, t_1, \dots, t_m$, утворюючи деякий процес, що складається із m етапів: $[t_0, t_1], \dots, [t_{m-1}, t_m]$. Стани S_i системи S у момент t_i , $i = \overline{1, m}$, опишемо деяким вектором

$$x^{(i)} = [x_1^{(i)}(t_i), x_2^{(i)}(t_i), \dots, x_n^{(i)}(t_i)]^T \in R^n.$$

Координати цього вектора називатимемо фазовими траєкторіями (або фазовими координатами). Сукупність станів характеризує процес переходу від першого стану $x^{(0)}$ до останнього $x^{(m)}$.

Наприклад, для системи, що описується рівнянням другого порядку

$$\frac{d^2 x(t)}{dt^2} = f(x, t),$$

стан системи визначається змінними стану x та dx/dt , тобто визначається функціями, що визначають рух точки x та її швидкість. Графіки цих функцій визначають так звані фазові траєкторії системи.

Передбачається, що на кожному етапі процес визначається як власною динамікою системи (станами), так і зовнішніми впливами

$$u^{(i)} = [u_1^{(i)}(t_i), u_2^{(i)}(t_i), \dots, u_r^{(i)}(t_i)]^T \in R^r; i = \overline{1, m}.$$

Назвемо ці впливи *керуваннями*. У результаті i -го етапу процес переходить у стан $x^{(i)}$, який визначається тільки станом $x^{(i-1)}$ та вибраним на цьому етапі керуванням $u^{(i)}$ і не залежить від "передісторії" процесу до i -го етапу. Зв'язок між станами можна описати рівняннями стану

$$x^{(i)} = f^{(i)}(x^{(i-1)}, x^{(i)}), \quad (1)$$

де $f^{(i)}(x, u)$, $i = \overline{1, m}$ -- деякі задані функції. Будемо вважати, що ефективність i -го стану оцінюється функцією мети $J^{(i)}(x^{(i-1)}, x^{(i)})$, а ефективність усього процесу -- адитивною функцією мети вигляду

$$J(x^{(0)}, \dots, x^{(m-1)}; u^{(1)}, \dots, u^{(m)}) = \sum_{i=1}^m J^{(i)}(x^{(i-1)}, u^{(i)}). \quad (2)$$

На фазову траєкторію і вибір керування можуть бути накладені обмеження вигляду

$$x^{(i)} \in X^{(i)}, i = \overline{1, m-1}, \quad (3)$$

$$u^{(i)} \in U^{(i)}(x^{(i-1)}), i = \overline{1, m}, \quad (4)$$

де $X^{(i)} \subset R^n$, $U^{(i-1)} \subset R^r$ -- задані множини. Функції (3) та (4), що не виходять за межі множин, називають допустими станами та керуваннями.

Окремо виділимо обмеження на початковий та кінцевий стани процесу:

$$x^{(0)} \in X^{(0)}, x^{(m)} \in X^{(m)}. \quad (5)$$

У більшості випадків множини станів $X^{(0)}$ і $X^{(m)}$ містять тільки по одній точці, тобто початкову і кінцеву траєкторії.

Задача оптимізації (або оптимального керування) описаним процесом ставиться так: серед усіх можливих (допустимих) керувань знайти таке керування $\{u^{(1)*}, \dots, u^{(m)*}\}$, яке за наявності умов (3), (4) переводить систему S зі стану $x^{(0)}$ у стан $x^{(m)}$ і для якого цільова функція (3) досягає свого екстремального значення.

Керування $\{u^{(1)*}, \dots, u^{(m)*}\}$ та відповідна до нього траєкторія $\{x^{(0)*}, \dots, x^{(m-1)*}\}$ називаються *оптимальними*.

Математична модель цієї задачі на максимум має вигляд:

$$\left. \begin{aligned} \max_u \quad & J(x^{(0)}, \dots, x^{(m-1)}; u^{(1)}, \dots, u^{(m)}), \\ x^{(i)} \quad &= f^{(i)}(x^{(i-1)}, u^{(i)}), \quad i = \overline{1, m}; \\ u^{(i)} \quad &\in U^{(i)}(x^{(i-1)}), \quad i = \overline{1, m}, \\ x^{(0)} \quad &\in X^{(0)}. \end{aligned} \right\} \quad (6)$$

Ця задача описує як дискретну, так і неперервну модель процесу оптимізації.

Пошук екстремуму функції мети J здійснюється поетапно, починаючи з останнього стану та керування.

Це стає можливим, коли функція мети є адитивна функція.

Адитивні функції і задачі

Визначення. Функція $f(x^{(0)}, \dots, x^{(N)})$ має назву адитивної, якщо вона має вигляд

$$f(x^{(0)}, \dots, x^{(N)}) = \sum_{i=0}^{N-1} f_i(x^{(i)}, x^{(i+1)}), \quad (7)$$

де $x^{(k)}$ -- вектори розміром n_k .

Це означає, що функція мети складається із суми функцій, що визначають стани системи на двох сусідніх значеннях цих станів. Для статичних процесів відома аналогічна теорема Колмогорова про подання функції багатьох змінних у вигляді суми адитивних функцій.

Якщо обмеження мають вигляд

$$x^{(i)} \in G_i, \quad i = \overline{0, N}, \quad (8)$$

задача пошуку мінімуму функції (7) також називається *адитивною*.

Адитивні задачі мають місце у багатьох галузях практичної діяльності людини. Вони мають наочну геометричну інтерпретацію, сутність якої полягає у такому. За припущення, що всі $n_k = n$, запровадимо у розгляд $(N+1)$ -вимірний простір пар $\{x, T\}$, в

якому побудовано гіперплощини Σ , що відповідають значенням $t = i, i = \overline{1, N}$. Тоді будь-яку сукупність векторів можна ототожнити з ламаною, що проходить у $(N+1)$ -вимірному просторі через точки $\{x^{(0)}, 0\} \in \Sigma_0, \dots, \{x^{(N)}, N\} \in \Sigma_N$. Її загальну довжину визначимо як суму довжин її ланцюгів, а довжину ланцюга, що сполучає точки $\{x^{(i)}, i\}, \{x^{(i+1)}, i+1\}$, будемо вимірювати величиною $f_i(x^{(i)}, x^{(i+1)})$. У цьому випадку довжина усієї ламаної є $f(x^{(0)}, \dots, x^{(N)})$.

Тепер початкову задачу можна сформулювати так: серед усіх допустимих $x^{(i)} \in G_i$ ламаних, що з'єднують гіперплощини Σ_0 та Σ_N , знайти ту, довжина якої є мінімальна.

У подальшому вектори $x^{(i)}$ будемо ототожнювати із точками $\{x_i, i\} \in \Sigma_i$. Множину усіх ламаних позначимо через Ω .

Розглянемо довільну точку $x_1 \in \Sigma_1$. Найкоротший відрізок, що з'єднує цю точку із гіперплощиною Σ_0 , має довжину

$$l(x_1) = \min_{x_0 \in G_0} f_0(x_0, x_1). \quad (9)$$

Оскільки

$$\min_{x_0 \in G_0} f(x_0, \dots, x_N) = l(x_1) + \sum_{i=1}^{N-1} f_i(x_i, x_{i+1}),$$

будь-яка ламана, що не містить відрізка $l(x_1)$, не може бути претендентом на те, щоб вона вважалася розв'язанням задачі мінімізації. Ці ламані утворюють множину ω_0 , яка й відкидається на першому кроці: $\Omega_1 \setminus \omega_0$. Якщо продовжити ці міркування, отримаємо найкоротшу ламану, що з'єднує точку x_{i+1} із гіперплощиною Σ_0 , якою буде

$$l(x_{i+1}) = \min_{x_i \in G_i} [l(x_i) + f_i(x_i, x_{i+1})]. \quad (10)$$

Ця формула являє собою рекурентне співвідношення, що описує багатокроковий процес пошуку розв'язання, яке приводить задачу пошуку мінімуму функції n змінних до послідовного пошуку мінімуму n функцій однієї змінної.

Структура адитивних задач є зручна і для застосування градієнтного спуску, оскільки

$$\frac{\partial f}{\partial x_i} = \frac{\partial f_i}{\partial x_i} + \frac{\partial f_{i-1}}{\partial x_i}.$$

Будемо говорити, що процес задовольняє принцип оптимальності, якщо будь-яка ділянка оптимальної траєкторії є оптимальна траєкторія.

6.1 Дискретні керуючі системи

Розглянемо динамічний процес, що протікає у дискретному часі. Позначимо стан процесу у момент t_i через $x(t_i)$, де x -- n -вимірний вектор і припустимо, що він описується рівнянням

$$x(t_{i+1}) = F_i[x(t_i), u(t_i)], \quad i = \overline{0, N-1}. \quad (11)$$

Початковий стан вважаємо заданим

$$x(t_0) = x_0. \quad (12)$$

Задача полягає у пошуку мінімуму функції вигляду

$$J = J(x_1, \dots, x_N; u_1, \dots, u_N) \quad (13)$$

із обмеженнями

$$x_i = x(t_i) \in X_i, \quad i = \overline{0, N}, \quad (14)$$

$$u_i = u(t_i) \in U_i, \quad i = \overline{0, N-1}. \quad (15)$$

Для того, щоб задача відшукування фазової траєкторії x_i та керування u_i , що реалізує мінімум функціоналу J_i приводилася до адитивного вигляду, дискретна система повинна мати низку властивостей. Насамперед, функціонал має бути адитивним, тобто

$$J = \sum_{i=0}^{N-1} f_i(x_i, x_{i+1}, u_i). \quad (16)$$

Крім того, система має допускати існування так званої *елементарної операції*, яка буде визначеною, якщо парі точок x_i, x_{i+1} можна співставити у відповідність керування u_i , що переводить систему за один такт із стану x_i у стан x_{i+1} . Тобто

$$u_i = B(x_i, x_{i+1}). \quad (17)$$

Тоді функція (16) прийме вигляд

$$J = \sum_{i=0}^{N-1} \tilde{f}_i(x_i, x_{i+1}), \quad (18)$$

де

$$\tilde{f}_i(x_i, x_{i+1}) = f_i(x_i, x_{i+1}; B(x_i, x_{i+1})).$$

Для того, щоб елементарна операція (17) могла бути побудована, необхідно, щоб система (11) мала розв'язання відносно u_i . Для цього розмірність вектора u_i має бути не менша за розмірність x_i .

Отже, якщо керування можна виключити із функціоналу, отримаємо адитивну задачу.

Інший напрямок дослідження дискретних керованих систем пов'язаний із виключенням фазових змінних.

Розглянемо для визначеності задачу із функцією мети

$$J = \sum_{i=1}^N F_i(x_i, u_{i-1}), \quad F_N = F_N(x_N) \quad (19)$$

та кінцево-різницевиими рівняннями

$$x_i = f_i(x_i, u_i), \quad i = \overline{0, N-1}. \quad (20)$$

Нехай, крім того, вектори x_i, u_i задовольняють умови (13)--(15). Оскільки початковий стан x_0 заданий, можна вважати, що $f_0 = f_0(u_0)$. Тоді з урахуванням рівняння (20) можна виключити вектори x_i із функціоналу (19):

$$J = F_1(u_0, f_0(u_0)) + F_2(u_1, f_1(u_1, f_0(u_0))) + \dots + F_N(f_{N-1}(u_{N-1}, f_{N-1}(u_{N-2}, f_{N-2} \dots)) \dots). \quad (21)$$

Функції такого вигляду звуться *функціями з послідовним виключенням змінних*.

6.1.1 Загальна схема динамічного програмування

Розглянемо останній, N -й крок процесу, що описується системою рівнянь (20) і припустимо, що система перебуває при $t = N - 1$ у деякому фіксованому стані x_{N-1} . Позначимо

$$J_k = \sum_{i=1}^k F_i(x_i, u_{i-1}). \quad (22)$$

Тоді

$$J = J_{N-1} + F_N(x_N).$$

Із урахуванням (186) подамо J у вигляді

$$J = J_{N-1} + F_N(f_{N-1}(x_{N-1}, u_{N-1})).$$

Якщо система уже перебуває у стані x_{N-1} , треба вибрати u_{N-1} так, щоб мінімізувати функцію

$$F_N(f_{N-1}(x_{N-1}, u_{N-1})).$$

Позначимо

$$S_{N-1}(x_{N-1}) = \min_{u_{N-1} \in U_{N-1}} F_N(f_{N-1}(x_{N-1}, u_{N-1})). \quad (23)$$

Ця рівність визначає деяку функцію розташування x_{N-1} . Величина

$$J = J_{N-1} + S_{N-1}(x_{N-1})$$

визначає те мінімальне значення функції J , яке може бути досягнуто, якщо траєкторія системи пройшла через точки $x_0, x_1, \dots, x_{N-1}$. У процесі обчислення величини $S_{N-1}(x_{N-1})$ відшукується керування u_{N-1} , що доставляє найменше значення функції (що задана)

$$F_N(f_{N-1}(x_{N-1}, u_{N-1})).$$

Отже, одночасно із скалярною функцією $S_{N-1}(x_{N-1})$ визначається й вектор- функція

$$u_{N-1} = \varphi_{N-1}(x_{N-1}), \quad (24)$$

яка ставить у відповідність кожному фазовому стану x_{N-1} керування u_{N-1} .

Фактично уже сформульовано правило відкидання не конкурентно- спроможних варіантів.

Розглянемо тепер другий крок. Функцію J можна тепер записати так:

$$\begin{aligned} J &= J_{N-2} + F_{N-1}(x_{N-1}, u_{N-1}) + S_{N-1} \\ &= J_{N-2} + F_{N-2}(f_{N-2}(x_{N-2}, u_{N-2})) + S_{N-1}(f_{N-2}(x_{N-2}, u_{N-2})). \end{aligned} \quad (25)$$

Якщо зафіксувати точку x_{N-2} , найменше значення функції J визначиться тільки керуванням u_{N-2} . Оскільки u_{N-2} входить тільки у дві останні складові в (191), то якщо знайти функцію

$$S_{N-2}(x_{N-2}) = \min_{u_{N-2} \in U_{N-2}} \{F_{N-1}(f_{N-2}(x_{N-2}, u_{N-2})) + S_{N-1}(f_{N-2}(x_{N-2}, u_{N-2}))\},$$

отримаємо мінімальне значення функції J :

$$J = J_{N-2} + S_{N-2}(x_{N-2}),$$

яке може бути досягнуто, коли траєкторія системи пройшла через точки $x_0, x_1, \dots, x_{N-2}$.

Одночасно із $S_{N-2}(x_{N-2})$ визначається також керування

$$u_{N-2} = \varphi_{N-2}(x_{N-2}).$$

Отже, якщо нам відомо, що у "момент часу" $t = N - 2$ система перебуває у стані x_{N-2} , для отримання найменшого значення функції J керування u_{N-2} , u_{N-1} на двох

останніх кроках слід обрати відповідно до формул

$$u_{N-2} = \varphi_{N-2}(x_{N-2});$$

$$u_{N-1} = \varphi_{N-1}(x_{N-1}) = \varphi_{N-1}[f_{N-1}(x_{N-2}, \varphi_{N-2}(x_{N-2}))].$$

На цьому кроці ми виключаємо із множини Ω_1 траєкторій, що залишилися, множину всіх траєкторій, для яких

$$u_{N-2} \neq \varphi_{N-2}(x_{N-2}).$$

Якщо процес продовжити, на кожному кроці визначимо функцію

$$S_{i-1}(x_{i-1}) = \min_{u_{i-1} \in U_{i-1}} \{F_i(f_{i-1}(x_{i-1}, u_{i-1}))\} \quad (26)$$

яка стану x_{i-1} системи за формулою

$$J = J_{i-1} + S_{i-1}(x_{i-1})$$

становить у відповідність те мінімальне значення J , яке може бути досягнуто, якщо траєкторія системи пройшла через точки $x_0, x_1, \dots, x_{i-1}$. Одночасно отримаємо функцію

$$u_{i-1} = \varphi_{i-1}(x_{i-1}).$$

Нарешті припустимо, що визначено функції $S_1(x_1)$ та $u_1 = \varphi_1(x_1)$; тоді

$$J = J_0 + S_1(x_1) = F_0(x_0) + S_1(x_1),$$

тож нам залишилося тільки визначити величину u_0 та мінімальне значення S_0 при заданому початковому стані x_0 :

$$S_0 = \min_{u_0 \in U_0} \{F_0(u_0) + S_1(x_0, u_0)\}.$$

$$u_0 = \varphi_0(x_0).$$

Отже, задачу вирішено. Число S_0 -- це мінімальне значення функції J на множині

$$U = U_0 \times U_1 \times \dots \times U_{N-1}.$$

Для реалізації цього значення S_0 треба побудувати послідовність векторів

$$u_0 = \varphi_0(x_0).$$

$$u_1 = \varphi_1(x_1) = \varphi_1[f_0(x_0, \varphi_0(u_0));$$

$$u_2 = \varphi_2(x_2) = \varphi_2[f_1(x_1, u_1)]$$

$$= \varphi_2[f_1(f_0(x_0, \varphi_0(x_0), \varphi_1(f_0(x_0), \varphi_0(x_0))))]; \dots$$

Рівняння (26), що визначає керування u_{i-1} , має назву *рівняння Беллмана*.

6.1.2 Числова реалізація

Для побудови числової схеми у просторі (x, t) побудуємо сітку із кроком Δx . Вузли цієї сітки позначимо через $P_k(i)$. На першому кроці процесу треба знайти функцію

$$S_{N-1}(P_k(N-1)) = \min_j F_N(f_{N-1}(P_k(N-1), u_{N-1}(j))).$$

На цьому ж кроці ми отримаємо також

$$u_{N-1}(k) = \varphi_{N-1}(P_k(N-1)).$$

На наступному кроці ми маємо побудувати таблицю для функції

$$S_{N-2}(P_k(N-2)) = \min_j \{F_{N-1}(f_{N-2}(P_k(N-2), u_{N-2}(j)))\}$$

$$+S_{N-1}(f_{N-2}(P_k(N-2), u_{N-2}(j)))\}.$$

Але, оскільки функція S_{N-1} задана таблично на сітці вузлів $P_i(N-1)$, серед цих вузлів може не виявитися вузла такого, щоб його координати визначалися компонентами вектора $(f_{N-2}(P_k(N-2), u_{N-2}(j)))$. Тобто точка x_{N-1} , яка відповідає керуванню $u_{N-2}(j)$, може не збігатися із жодним вузлом у гіперплощині Σ_{N-1} , а, значить, вона відсутня у таблиці значень функції S_{N-1} .

Перехід із заданого стану x_{N-2} у заданий стан x_{N-1} можливий тільки тоді, коли існує керування u_{N-2} , яке є корінь векторного рівняння

$$x_{N-1} = f_{N-1}(x_{N-2}, u_{N-2}). \quad (27)$$

Для цього необхідно, щоб розмірність вектора u була не менше за розмірність вектору x . У цьому випадку за фіксованого x_{N-2} формула (27) дає деяке відображення U_{N-2} на Σ_{N-2} . Позначимо його через $Q(U_{N-2}, x_{N-2})$. Якщо деякий вузол $P_j(N-1) \in Q(U_{N-2}, x_{N-2})$, говорять, що він є досяжний із точки x_{N-2} .

Елементарною операцією (стосовно даного випадку) називають процедуру відшукування дійсного кореня рівняння (193), тобто функції u_{N-2} , що залежить від x_{N-1} та x_{N-2} :

$$u_{N-2} = \varphi_{N-2}^*(x_{N-1}, x_{N-2}).$$

Тепер можна подати функцію S_{N-2} у такій формі:

$$S_{N-2}(P_k(N-2)) = \min_j \{F_{N-1}[f_{N-2}(P_k(N-2), \varphi_{N-2}^*(P_j(N-1), P_k(N-2)))] \\ + S_{N-1}[f_{N-2}(P_k(N-2), \varphi_{N-2}^*(P_j(N-1), P_k(N-2)))]\}.$$

Для побудови таблиці функції $S_{N-2}(P_k(N-2))$ досить перебрати значення вузлів $P_j(N-1)$. Відмітимо, що досить заповнити таблицю керування $u_{N-2}(P_k(N-2))$.

Якщо розмірність вектора u менша за розмірність вектору x , ситуація суттєво ускладнюється, оскільки розмірність множини досяжності у загальному випадку виявляється меншою за розмірність простору Σ_{N-1} . У цьому випадку треба модифікувати алгоритм побудови.

Кожний із вузлів $P_j(i)$ оточимо деякою множиною $R_j(i)$, тобто вузол $P_j(i)$ є центр деякого паралелепіпеду з довжиною ребра 2ε . Будемо відносити до числа вузлів, що є недосяжні із точки x_{i-1} , усі вузли, в околі яких $R_j(i)$ мають із множиною досяжності спільні точки.

Отже, нехай функцію $S_i(x_i)$ уже побудовано. Це означає, що ми маємо таблицю значень $S_i(P_j(i))$ та керування $u_i(x_i)$. Запишемо функції $S_i(x_i)$ та $F_i(x_i, u_i)$ у такому вигляді:

$$S_i(x_i) = S_i[f_{i-1}(x_{i-1}, u_{i-1}(j_k))]; \\ F_i(P_k(i)) = F_i[f_{i-1}(x_{i-1}, u_{i-1}(j_k))]. \quad (28)$$

За цими таблицями ми легко будемо таблицю для функції

$$S_{i-1}(x_{i-1}) = \min_k \{F_i[f_{i-1}(x_{i-1}, u_{i-1}(j_k))] + S_i[f_{i-1}(x_{i-1}, u_{i-1}(j_k))]\}. \quad (29)$$

Операція (29) кожному вузлу $P_k(i-1)$, що належить площині Σ_{i-1} , співставляє у відповідність єдине значення керування $u_{i-1}(j_k)$. Цей процес надає можливість отримати деяку послідовність керувань $u_0, u_1, \dots, u_{N-1}$ та за допомогою формули (20) побудувати фазову траєкторію $x_0, x_1, \dots, x_N$, й отже, обчислити мінімальне значення J .

Викладена процедура надає можливість обчислити мінімальне значення J тим точніше, чим на густішій сітці виконуються обчислення. У зв'язку з цим для розв'язання подібних задач виникає потреба у використанні ітеративних методів.

6.1.3 Приклади

Приклад 1. У визначений момент часу на підприємстві встановлено нове обладнання. Залежність продуктивності цього обладнання від часу його використання, а також залежність витрат на утримання і ремонт обладнання відомі і наведені у табл. 6.1. $R(\tau)$ -- річний випуск продукції; $Z(\tau)$ -- річні витрати, що пов'язані із утриманням та ремонтом обладнання.

Табл. 6.1 Залежність продуктивності цього обладнання від часу

Роки	0	1	2	3	4	5
$R(\tau)$ (у.о.)	80	75	65	60	60	55
$Z(\tau)$ (у.о.)	20	25	30	35	40	55

Відомо, що витрати, пов'язані із придбанням і встановленням обладнання, що замінюється, списується.

Треба скласти такий план заміни обладнання упродовж п'яти років, за якого загальний прибуток за даний період часу був би максимальним.

Розв'язання.

Це завдання можна розглядати як завдання динамічного програмування, в якому в якості системи S є обладнання. Стани цієї системи визначаються фактичним часом використання обладнання (його віком) τ , тобто описуються одним параметром -- τ . Керуваннями є рішення про заміну або збереження обладнання, які приймаються на початку кожного року.

Позначимо u_1 -- рішення про збереження обладнання, а u_2 -- рішення про заміну обладнання. Тоді завдання полягатиме у відшуванні такої стратегії керування, яка визначається рішеннями, що приймаються до початку кожного року, за якої загальний прибуток підприємства за п'ять років буде максимальний.

Отже, сформульоване вхідне завдання у термінах завдання динамічного програмування. Воно характеризується властивостями адитивності і відсутністю післядії. Розв'язання задачі можна знайти за допомогою розв'язання задачі динамічного програмування, що реалізується у два етапи. На першому етапі, рухаючись від початку п'ятого року до початку першого року для кожного допустимого стану обладнання знайдемо умовне оптимальне керування (рішення), а на другому, рухаючись від початку першого року до початку п'ятого року із умовних оптимальних рішень для кожного року оберемо оптимальний план заміни обладнання.

Для визначення умовних оптимальних рішень спочатку необхідно скласти функціональне рівняння Беллмана. Оскільки ми припустили, що до початку k -го року ($k = \overline{1,5}$) може прийматися тільки одне з двох рішень -- замінювати чи не замінювати

обладнання, прибуток підприємства за k -й рік складає

$$F_k(\tau^{(k)}, u_k) = \begin{cases} R(\tau^{(k)}) - Z(\tau^{(k)}), & \text{їдє} u_1, \\ R(\tau^{(k)} = 0) - Z(\tau^{(k)} = 0) - C_n, & \text{їдє} u_2, \end{cases}$$

де $\tau^{(k)}$ -- вік обладнання на початок k -го року; u_k -- керування, яке реалізується на початок k -го року; C_n -- коштовність нового обладнання.

Отже, у даному випадку рівняння Беллмана має вигляд

$$F_k(\tau^{(k)}) = \max_{\tau} \begin{cases} R(\tau^{(k)}) - Z(\tau^{(k)}) + F_{k+1}(\tau^{(k+1)}), \\ R(\tau^{(k)} = 0) - Z(\tau^{(k)} = 0) - C_n + F_{k+1}(\tau^{(k+1)} = 1), \end{cases} \quad (30)$$

Почнемо із визначення умовно оптимального керування (рішення) для останнього (п'ятого) року, для чого знаходимо множину допустимих станів обладнання на початок даного року.

Використовуючи рівняння Беллмана, переходимо до розв'язання початкової задачі. Почнемо з визначення умовного оптимального керування для останнього (п'ятого) року, для чого знаходимо множину допустимих станів обладнання на початок даного року. Оскільки у початковий момент у наявності є нове обладнання ($\tau^{(1)} = 0$), вік обладнання на початок п'ятого року може складати 1, 2, 3 і 4 роки. Тому допустимі стани системи на даний період часу такі: $\tau^{(5)} = 1$, $\tau^{(5)} = 2$, $\tau^{(5)} = 3$, $\tau^{(5)} = 4$. Для кожного з цих станів знайдемо умовно оптимальний розв'язок і відповідне до нього значення функцій $F_5(\tau^{(5)})$.

Із (Ошибка! Источник ссылки не найден.) і співвідношення $F_5(\tau^{(k+1)}) = 0$ (розглядається останній рік розрахункового періоду). отримаємо

$$F_5(\tau^{(5)}) = \max_{\tau} \begin{cases} R(\tau^{(5)}) - Z(\tau^{(5)}), \\ R(\tau^{(5)} = 0) - Z(\tau^{(5)} = 0) - C_n, \end{cases} \quad (31)$$

Із урахуванням $\tau^{(5)} = 1$, а також даних, що наведено у табл. 6.1, отримуємо:

$$F_5(\tau_1^{(5)}) = \max \begin{cases} R(\tau^{(5)} = 1) - Z(\tau^{(5)} = 1), \\ R(\tau^{(5)} = 0) - Z(\tau^{(5)} = 0) - C_n, \end{cases} = \max \begin{cases} 75 & -25, \\ 80 & -20 - 40, \end{cases} = 50, \\ u^o = u_1.$$

Виконаємо аналогічні обчислення для інших допустимих станів обладнання на початок п'ятого року:

$$F_5(\tau_2^{(5)}) = \max \begin{cases} 65 & -30, \\ 80 & -20 - 40, \end{cases} = 35, \quad u^o = u_1. \\ F_5(\tau_3^{(5)}) = \max \begin{cases} 60 & -35, \\ 80 & -20 - 40, \end{cases} = 25, \quad u^o = u_1. \\ F_5(\tau_4^{(5)}) = \max \begin{cases} 65 & -45, \\ 80 & -20 - 40, \end{cases} = 20, \quad u^o = u_2.$$

Табл.6.2 Умовно оптимальний розв'язок для $\tau^{(5)}$

Вік обл. $\tau^{(5)}$	$F_5(\tau^{(5)})$ (y.o.)	u^o
1	50	u_1

2	35	u_1
3	25	u_1
4	20	u_2

Розглянемо тепер можливі стани на початок 4-го року. Вочевидь, допустимими станами є $\tau_1^{(4)} = 1$, $\tau_2^{(4)} = 2$, $\tau_3^{(4)} = 3$. Для кожного з них визначаємо умовно оптимальний розв'язок і відповідне до нього значення функції $F_4(\tau^{(4)})$. Для цього скористаємося рівнянням (30) і табл. 6.1, 6.2 та отриманими даними. Зокрема, для $\tau_1^{(4)} = 1$ маємо:

$$F_4(\tau_1^{(4)}) = \max \begin{cases} R(\tau^{(4)} = 1) - Z(\tau^{(4)} = 1) + F_5(\tau^{(5)} = 2), \\ R(\tau^{(4)} = 0) - Z(\tau^{(4)} = 0) - C_n + F_5(\tau^{(5)} = 1), \end{cases}$$

$$= \max \begin{cases} 75 & -25 + 35, \\ 80 & -20 - 40 + 50, \end{cases} = 85, \quad u^o = u_1.$$

$$F_4(\tau_2^{(4)}) = \max \begin{cases} 65 & -25 + 35, \\ 80 & -20 - 40 + 50, \end{cases} = 75, \quad u^o = u_1.$$

$$F_4(\tau_3^{(4)}) = \max \begin{cases} 65 & -30 + 20, \\ 80 & -20 - 40 + 50, \end{cases} = 70, \quad u^o = u_2.$$

Результати обчислень запишемо у вигляді табл. 6.3.

Табл. 6.3. Умовно оптимальний розв'язок для $\tau^{(4)}$

Вік обл. $\tau^{(4)}$	$F_5(\tau^{(4)})$ (y.o.)	u^o
1	85	u_1
2	70	u_1
3	70	u_2

Визначимо тепер умовно оптимальний розв'язок для кожного з допустимих станів обладнання на початок третього року. Вочевидь, допустимими станами є $\tau_1^{(3)} = 1$, $\tau_2^{(3)} = 2$.

$$F_3(\tau_1^{(3)}) = \max \begin{cases} R(\tau^{(3)} = 1) - Z(\tau^{(3)} = 1) + F_4(\tau^{(4)} = 2), \\ R(\tau^{(3)} = 0) - Z(\tau^{(3)} = 0) - C_n + F_4(\tau^{(4)} = 1), \end{cases}$$

$$= \max \begin{cases} 75 & -25 + 70, \\ 80 & -20 - 40 + 85, \end{cases} = 120, \quad u^o = u_1.$$

$$F_3(\tau_2^{(3)}) = \max \begin{cases} R(\tau^{(3)} = 2) - Z(\tau^{(3)} = 2) + F_4(\tau^{(4)} = 3), \\ R(\tau^{(3)} = 0) - Z(\tau^{(3)} = 0) - C_n + F_4(\tau^{(4)} = 1), \end{cases}$$

$$= \max \begin{cases} 60 & -35 + 70, \\ 80 & -20 - 40 + 85, \end{cases} = 105, \quad u^o = u_2.$$

Табл. 6.4. Умовно оптимальний розв'язок для $\tau^{(3)}$

Вік обл. $\tau^{(3)}$	$F_5(\tau^{(3)})$ (y.o.)	u^o
1	120	u_1
2	105	u_2

Залишилося розглянути допустимі стани обладнання на початок другого року. Вочевидь, на даний момент часу вік обладнання може дорівнювати одному року, тому треба порівняти тільки два можливі варіанти: зберегти обладнання або виконати заміну. Маємо $F_2(\tau^{(2)}) = 120$, $u^0 = u_1$.

Відповідно до умови у початковий момент встановлено нове обладнання ($\tau_1^{(1)} = 0$), тому проблеми вибору між збереженням і заміною обладнання не існує: обладнання слід зберегти. Отже, умовно оптимальним розв'язком є u_1 , а значення функції

$$F_1(\tau_1^{(1)}) = R(\tau_2^{(1)} = 0) - Z(\tau_1^{(1)} = 0) + F_2(\tau_1^{(1)} = 1) = 80 - 20 + 155 = 215.$$

Отже, максимальний прибуток підприємства може дорівнювати 215 у.о. Він відповідає оптимальному плану заміни обладнання, який дістаємо на основі даних таблиць 6.1—6.4, тобто внаслідок реалізації другого етапу обчислювального процесу, що полягає у проходженні всіх розглянутих кроків від початку першого до початку п'ятого років.

Для першого кроку розв'язання безальтернативне: слід зберегти обладнання. Тож вік обладнання на початок другого року дорівнює одному року. Тоді оптимальним розв'язком для другого року є рішення про збереження обладнання. Реалізація такого рішення призводить до того, що вік обладнання на початок третього року становитиме два роки. За такого віку (3) обладнання у третьому році слід замінити. Вік обладнання на початок п'ятого року складе два роки, тобто змінювати обладнання на початок п'ятого року недоцільно.

Отже, знайдено оптимальний план заміни обладнання. У перший, другий, четвертий та п'ятий роки слід зберегти наявне обладнання, а на початку третього року -- виконати заміну обладнання.

Приклад 2.

У розпорядженні міністерства, якому підпорядковуються k підприємств, виділені кошти у розмірі K у.о. для використання їх на розвиток підприємств упродовж m років. Ці кошти на початку кожного року (тобто у моменти $t_1, t_2, \dots, t_m$) розподілено між підприємствами. Одночасно з цим між підприємствами розподіляється прибуток, отриманий ними за минулий рік. Таким чином, на початку кожного i -го року j -те підприємство отримує у своє розпорядження $x_i^{(j)}$ у.о.

Визначити такі значення $x_i^{(j)}$, тобто знайти такі розподіли коштів і прибутку між підприємствами, за яких за m років забезпечується отримання максимального прибутку всіма підприємствами.

Розв'язання.

Нехай j -му підприємству на i -й рік виділяється $x_i^{(j)}$ у.о. Будемо розглядати цей розподіл коштів як реалізацію деякого керування u_i . Керування полягає у тому, що на i -му кроці першому підприємству виділяється $x_i^{(1)}$ у.о., другому -- $x_i^{(2)}$ тощо. Сукупність чисел $x_i^{(j)}$, $j = \overline{1, k}$, визначає усю сукупність керувань $u_1, u_2, \dots, u_m$ на m кроках розподілення коштів як m точок у k -вимірному просторі.

Критерієм оцінки якості вибраного розподілу коштів, тобто керувань, що реалізуються, обрано сумарний прибуток за m років, який залежить від усієї сукупності керувань: $W = W(u_1, u_2, \dots, u_m)$. Отже, завдання полягає у виборі таких керуючих впливів u_i^* (тобто у такому розподілі коштів), за яких функція W набуває максимальне значення.

Бачимо, що це завдання є багатоетапне. Ця багатоетапність диктується її умовами, що передбачають прийняття рішень на початку кожного року періоду, що розглядається. Разом з тим, у низці інших завдань динамічного програмування багатоетапність безпосередньо не впливає із умов.

Наведемо приклад і дамо формулювання завдання у термінах загального завдання динамічного програмування.

Нехай $S = 7$ млн. у.о., $n = 3$, а значення X_i і $f(X_i)$ наведено у табл. 6.5, $f(X_i)$ -- приріст випуску продукції залежно від обсягів капіталовкладень X_i .

Для розв'язання даного завдання динамічного програмування слід скласти рекурентне співвідношення Беллмана. У даному випадку це співвідношення призводить до таких функціональних рівнянь:

$$\begin{aligned}\varphi_1(X) &= \max_{0 \leq X_1 \leq X} \{f_1(X_1)\}; \\ \varphi_2(X) &= \max_{0 \leq X_2 \leq X} \{f_1(X_2) + \varphi_1(X - X_2)\}; \dots; \\ \varphi_{n-1}(X) &= \max_{0 \leq X_{n-1} \leq X} \{f_{n-1}(X_{n-1}) + \varphi_{n-2}(X - X_{n-1})\}.\end{aligned}\tag{32}$$

Тут функції $\varphi_i(X)$, $i = 1, n-1$, визначають максимальний приріст випуску продукції за відповідних розподілів X у.о. капіталовкладень між i підприємствами. Тому значення функції $\varphi_n(X)$ обчислюється лише для одного значення $X = S$, оскільки обсяг капіталовкладень, які виділяються для усіх n підприємств, дорівнює S у.о.

Табл. 6.5. Вхідні дані

X_i (у.о.)	$f_1(X_1)$	$f_2(X_2)$	$f_3(X_3)$
0	0	0	0
100	30	50	40
200	50	80	50
300	90	90	110
400	110	150	120
500	170	190	180
600	180	210	220
700	210	220	240

Згідно із (32) і даними табл. 6.5 шукатимемо спочатку умовно оптимальні, а потім оптимальні розподіли капіталовкладень між підприємствами.

Почнемо із визначення умовно оптимальних капіталовкладень, що виділяються на розвиток першого підприємства. Для цього знаходимо значення $\varphi_1(X)$ для кожного

$$X = \{0, 100, 200, 300, 400, 500, 600, 700\}.\tag{33}$$

Нехай $X = 0$, тоді $\varphi_1(0) = 0$. Візьмемо тепер $X = 100$. Тоді за даними табл. 6.5 дістанемо

$$\varphi_1(100) = \max\{0, 30\} = 30, \quad X_1^o = 100.$$

Аналогічно знаходимо умовно оптимальні розв'язки для інших значень X :

$$\varphi_1(200) = \max\{0, 30, 50\} = 50, \quad X_1^o = 200;$$

$$\varphi_1(300) = \max\{0, 30, 50, 90\} = 90, \quad X_1^o = 300;$$

$$\varphi_1(400) = \max\{0, 30, 50, 90, 110\} = 110, \quad X_1^o = 400;$$

$$\varphi_1(500) = \max\{0, 30, 50, 90, 110, 170\} = 170, \quad X_1^o = 500;$$

$$\varphi_1(600) = \max\{0, 30, 50, 90, 110, 170, 180\} = 180, \quad X_1^o = 600;$$

$$\varphi_1(700) = \max\{0, 30, 50, 90, 110, 170, 180, 210\} = 210, \quad X_1^o = 700;$$

Результати обчислень відповідні умовно оптимальні розв'язки записуємо у табл.

6.6

Табл. 6.6. Умовно оптимальний розв'язок

X_1 (тис.у.о.)	$\varphi_1(X)$	X_1^o
0	0	0
100	30	100
200	50	200
300	90	300
400	110	400
500	170	500
600	180	600
700	210	700

Використаємо тепер дані табл. 6.5, 6.6. Умовно оптимальні обсяги капіталовкладень, що виділяються другому підприємству. Знайдемо

$$\varphi_2(X) = \max_{0 \leq X_2 \leq X} \{f_2(X_2) + \varphi_2(X - X_2)\}$$

для кожного з допустимих значень X , б що дорівнюють (33):

$$\varphi_2(X) = 0, \quad X_2^o = 0;$$

$$\varphi_2(100) = \max\{0 + 30; 50 + 0\} = 50; \quad X_2^o = 100;$$

$$\varphi_2(200) = \max\{0 + 50; 50 + 30; 80 + 0\} = 80; \quad X_2^o = 100;$$

$$\varphi_2(300) = \max\{0 + 90; 50 + 50; 80 + 30; 90 + 0\} = 100; \quad X_2^o = 200;$$

$$\varphi_2(400) = \max\{0 + 110; 50 + 90; 80 + 50; 150 + 0\} = 150; \quad X_2^o = 400;$$

$$\varphi_2(500) = \max\{0 + 170; 50 + 110; 80 + 90; 150 + 30; 190 + 0\} = 190; \quad X_2^o = 500;$$

$$\varphi_2(600) = \max\{0 + 180; 50 + 170; 80 + 110; 150 + 50; 190 + 30; 210 + 0\} = 220; \quad X_2^o = 100;$$

$$\varphi_2(700) = \max\{0 + 210; 50 + 180; 80 + 170; 150 + 90; 190 + 50; 210 + 30; 220 + 0\} = 250;$$

$$X_2^o = 200.$$

Отримані умовно оптимальні обсяги капіталовкладень, що призначаються другому підприємству, записуємо у вигляді табл. **Ошибка! Источник ссылки не найден..** Тут X -- обсяги капіталовкладень, що призначаються першому і другому підприємствам; $\varphi_2(X)$ -- максимальний приріст випуску продукції; X_2^o -- умовно оптимальний обсяг капіталовкладень, що виділяється другому підприємству.

Переходимо до відшукування значень

$$\varphi_3(X) = \max_{0 \leq X_3 \leq X} \{f_3(X_3) + \varphi_3(X - X_3)\}$$

із використанням даних табл. 6.5, 6.7

Табл. 6.7. Умовно оптимальний розв'язок

X (тис. у.о.)	$\varphi_2(X)$	X_2^o
0	0	0
100	50	100
200	80	100
300	110	200
400	150	400
500	190	500
600	220	100
700	250	200

Оскільки у даному випадку число підприємств дорівнює трьом, виконуємо обчислення лише для одного значення $X = 700$.

$$\varphi_3(700) = \max\{0 + 50; 40 + 220; 50 + 190; 110 + 150; 120 + 110; 180 + 80; 220 + 50; 240 + 0\} = 270;$$

$$X_3^o = 600.$$

Отже, максимальний приріст випуску продукції складає 270 тис. у.о. Це станеться тоді, коли 3-му підприємству призначається 600 тис. у.о., а першому і другому -- 100 тис. у.о.

6.2 Метод ДП для неперервних систем

6.2.1 Автономна система

Нехай рух об'єкту описується рівнянням

$$\frac{dx}{dt} = f(x, u), \quad (34)$$

$x = [x_1, x_2, \dots, x_n]$ -- n -вимірний вектор стану; $u = [u_1, u_2, \dots, u_m] \in U$ -- m -вимірний вектор управління; f -- n -вимірний вектор. Функціонал

$$I = \int_0^T G(x, u) dt. \quad (35)$$

Вважаємо фіксованими початковий стан x і кінцевий стан x^* . Час переходу з початкового стану у кінцевий не фіксується. Оскільки мета оптимізації полягає в отриманні оптимальної синтезуючої функції (оптимальної стратегії), початковою точкою x може бути будь-яка точка фазового простору.

Мінімальне значення функціоналу однозначно визначається початковим значенням x . Позначимо мінімальне значення функціоналу $S(x) = S(x_1, x_2, \dots, x_n)$.

Нехай $x(t)$, $0 \leq t \leq T$ -- оптимальна траєкторія, що переводить фазову точку із початкового положення $x(0)$ у кінцеву точку x^* . Тоді

$$S(x) = \min_{u(t) \in U} \int_0^T G(x(t), u(t)) dt. \quad (36)$$

Подамо функціонал у вигляді

$$\int_0^T G(x(t), u(t)) dt = \int_0^\Delta G(x(t), u(t)) dt + \int_\Delta^T G(x(t), u(t)) dt. \quad (37)$$

Нехай в інтервалі $(0, \Delta)$ вибрано управління $u(t)$, а потім відповідно до принципу оптимальності вибирається оптимальне управління. Тоді

$$S(x(\Delta)) = \int_\Delta^T G(x(t), u(t)) dt. \quad (38)$$

У силу неперервності траєкторії $x(t)$ $x(\Delta) = x(0) + x(0) \cdot \Delta + j(\Delta)$.

Із урахуванням рівняння (34) маємо

$$x(\Delta) = x + f(x, u)|_{t=0} \cdot \Delta + o(\Delta).$$

Отже,

$$\begin{aligned} \int_\Delta^T G(x(t), u(t)) dt &= S(x + f(x, u) \cdot \Delta + o(\Delta)). \\ \int_0^\Delta G(x(t), u(t)) dt &= G(x, u) \cdot \Delta + o(\Delta). \end{aligned}$$

Функціонал (37) набуває значення

$$S(x) = \min_{u(t) \in U} [G(x, u) \cdot \Delta + o(\Delta) + S(x + f(x, u) \cdot \Delta + o(\Delta))]. \quad (39)$$

Оскільки функція $S(x)$ припускається неперервно диференційовною,

$$S(x + f(x, u) \cdot \Delta + o(\Delta)) = S(x) + \frac{dS}{dx} f(x, u) \cdot \Delta + o(\Delta).$$

За правилами диференціювання скалярної функції по векторному аргументу

$$\frac{dS}{dx} = \left(\frac{\partial S}{\partial x_1}, \frac{\partial S}{\partial x_2}, \dots, \frac{\partial S}{\partial x_n} \right).$$

Із (Ошибка! Источник ссылки не найден.) отримуємо

$$0 = \min_{u \in U} \left[G(x, u) + \frac{dS}{dx} f(x, u) \right] \Delta + o(\Delta). \quad (40)$$

Здійснимо граничний перехід за $\Delta \rightarrow 0$:

$$0 = \min_{u \in U} \left[G(x, u) + \frac{dS}{dx} f(x, u) \right]. \quad (41)$$

Це -- функціональне рівняння Беллмана. До нього треба приєднати межу умову

$$S(x^*) = 0. \quad (42)$$

Розв'язання задачі (41), (42) разом із функцією $S(x)$, яка задає залежно від початкової точки x мінімальне значення функціоналу, визначається також функція $u(x)$, що задає оптимальну траєкторію, або оптимальну синтезуючу функцію.

Рівняння Беллмана задає необхідну умову мінімуму.

Приклад. Розглянемо об'єкт, рух якого задається рівнянням

$$\frac{dx_1}{dt} = x_2, \quad \frac{dx_2}{dt} = u.$$

В якості кінцевої точки x^* виберемо початок координат, тобто $x^* = 0$. Якість процесу управління будемо оцінювати функціоналом

$$I = \int_0^T [x_1^2 + x_2^2 + u^2] dt. \quad (43)$$

Отже, мова йде про визначення оптимальної стратегії $u = f(x_1, x_2)$, яка забезпечить переведення фазової точки із довільного початкового стану у початок координат, й при тому так, щоб на траєкторії руху функціонал (43) набував найменше значення.

Запишемо функціональне рівняння Беллмана

$$0 = \min_u \left[x_1^2 + x_2^2 + u^2 + \frac{\partial S}{\partial x_1} x_2 + \frac{\partial S}{\partial x_2} u \right]. \quad (44)$$

Оскільки на параметр u відсутні будь-які обмеження, для визначення мінімуму необхідно продиференціювати рівняння (Ошибка! Источник ссылки не найден.) по u :

$$2u + \frac{\partial S}{\partial x_2} \rightarrow u = -\frac{1}{2} \frac{\partial S}{\partial x_2}. \quad (45)$$

Рівняння Беллмана набуває вигляду

$$x_1^2 + x_2^2 + \frac{\partial S}{\partial x_2} - \frac{1}{4} \left(\frac{\partial S}{\partial x_2} \right)^2 = 0. \quad (46)$$

Будемо шукати розв'язок у вигляді квадратичної форми

$$S(x) = C_1 x_1^2 + C_2 x_1 x_2 + C_3 x_2^2.$$

Тоді

$$\frac{\partial S}{\partial x_1} = 2C_1x_1 + C_2x_2; \quad \frac{\partial S}{\partial x_2} = C_2x_1 + 2C_3x_2.$$

Маємо

$$x_1^2 + x_2^2 - \frac{1}{4}(4C_3^2x_2^2 + 4C_2x_1x_2 + C_3^2x_1^2) + 2C_1x_1x_2 + C_2x_2^2 = 0.$$

Коефіцієнти C_1, C_2, C_3 визначаються із системи рівнянь

$$\begin{cases} 1 - \frac{1}{4}C_2^2 &= 0, \\ 1 - C_3^2 + C_3 &= 0, \\ 2C_1 - C_2C_3 &= 0. \end{cases} \quad (47)$$

Ця система має два розв'язки:

$$C_2 = 2, C_3 = \sqrt{3}, C_1 = \sqrt{3};$$

$$C_2 = 2, C_3 = -\sqrt{3}, C_1 = -\sqrt{3}.$$

Згідно з (45) маємо дві синтезуючі функції:

$$u = -x_1 - \sqrt{3}x_2, \quad u = -x_1 + \sqrt{3}x_2.$$

Отримуємо дві лінійні системи, причому лінійна система, що породжена функцією $u = -x_1 + \sqrt{3}x_2$, виявляється нестійкою й отже не може забезпечити перехід фазової точки у початок координат.

6.2.2 Неавтономна система

Розглянемо неавтономну систему рівнянь

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_m, t), \quad i = \overline{1, n}. \quad (48)$$

Функціонал

$$I = \int_{t_0}^T G(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_m, t) dt. \quad (49)$$

Розглядається задача про відшукування управління, що задане у вигляді оптимальної стратегії, яку здійснює переведення фазової точки $x = (x_1, x_2, \dots, x_n)$ системи (48) із заданого, але будь-якого початкового стану у деяку задану точку $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ так, щоб функціонал (49) набував мінімальне значення. Початковий момент часу t_0 вважається фіксованим, але може бути будь-якою величиною. Кінцевий момент часу T вважатимемо не заданим, а визначається із умови проходження траєкторії $x(t)$, $t_0 \leq t \leq T$, через точку x^* , тобто $x(T) = x^*$. Скористаємося способом приведення неавтономної задачі оптимального управління до автономної. Запишемо рівняння (48) у векторній формі

$$\frac{dx}{dt} = f(x, u, t), \quad (50)$$

Приєднаємо до рівнянь (48) рівняння

$$\frac{dx_{n+1}}{dt} = 1, \rightarrow x_{n+1} = t; \quad x_{n+1}(t_0) = t_0. \quad (51)$$

Маємо, таким чином, автономну систему диференціальних рівнянь

$$\frac{dx}{dt} = f(x, u, t), \quad \frac{dx_{n+1}}{dt} = 1. \quad (52)$$

і функціонал вигляду

$$I = \int_{t_0}^T G(x, u, x_{n+1}) dt. \quad (53)$$

Задача оптимального управління (52), (53) є задача із закріпленим лівим і рухомим правим кінцями траєкторії. Функціональне рівняння Беллмана має вигляд

$$0 = \min_{u \in U} \left[G(x, u, x_{n+1}) + \frac{\partial S}{\partial x} f(x, u, x_{n+1}) + \frac{\partial S}{\partial x_{n+1}} \right]. \quad (54)$$

Оскільки $x_{n+1} = t$, рівняння (**Ошибка! Источник ссылки не найден.**) можна записати у вигляді

$$-\frac{\partial S}{\partial t} = \min_{u \in U} \left[G(x, u, x_{n+1}) + \frac{\partial S}{\partial x} f(x, u, x_{n+1}) \right]. \quad (55)$$

Межова умова

$$S(x^*, T) = 0. \quad (56)$$

6.3 Контрольні запитання

1. Що є предметом динамічного програмування?
2. Сформулюйте постановку задачі динамічного програмування у загальному вигляді.
3. Сформулюйте принцип оптимальності Беллмана.
4. Чому задачі динамічного програмування називають багатокроковими?
5. З яких етапів складається алгоритм розв'язання задачі динамічного програмування?
6. Яким чином під час розв'язання задачі динамічного програмування проводиться умовна оптимізація?
7. Яким чином під час розв'язання задачі динамічного програмування проводиться безумовна оптимізація?
7. Перелічіть класи задач, розв'язуваних за допомогою динамічного програмування.
8. Сформулюйте постановку задачі розподілу ресурсів методом динамічного програмування.
9. Охарактеризуйте основні етапи алгоритму розв'язання задачі розподілу ресурсів методом динамічного програмування.
10. У чому полягають переваги динамічного програмування?

Моделі біологічних та медичних процесів

Розділ 7 Математичне моделювання імунних систем

7.1 Вступ

В якості одного з найефективніших підходів до побудови оптимальної стратегії з вивчення та прогнозування вірусних захворювань є математичне моделювання, що надає можливість аналізувати вплив різних факторів та їхніх сполучень на індивідуальному і популяційному рівнях. Прикладом моделей такого типу можуть слугувати математичні моделі прогнозування, що описують взаємозв'язок між показниками здоров'я людини та факторами навколишнього середовища.

З огляду на складність взаємодії складових імунної системи запроваджено низку спрощених припущень.

Популяції клітин і вірусів вважаються рівномірно розподіленими по епітеліальному шару органу-мішені у будь-який момент часу. Швидкість змінювання змінних визначається поточним значенням усіх змінних. Вважається, що основні процеси регуляції імунного захисту здійснюються у трьох локальних об'ємах: головному мозку (гіпофізі, гіпоталамусі), черевній порожнині та об'ємі органу-мішені. Взаємодія між цими об'ємами здійснюється із запізненням у часі.

Запуск захисних механізмів здійснюється після взаємодії макрофагів із клітинами органу-мішені, що відмерли, які зруйновано внаслідок життєвого циклу вірусу. Видалення макрофагами клітин, уражених вірусом, супроводжується синтезом інформаційних молекул (цитокіну) інтерлейкіну-1 [22].

Поява у крові підвищеного вмісту інтерлейкіну-1 сприяє виробленню інтерлейкіну-2 Т-хелперами і стимулює специфічні рецептори гіпоталамусу на вироблення рілізінг-гормону кортиколіберину.

Метою регуляторного впливу інтерлейкіну-2 є НК-клітини, цитотоксичні Т-лімфоцити і В-клітини [15,32,35]. Основна функція НК-клітин пов'язана із знищенням інфікованих клітин на ранніх етапах захисту організму проти вірусних інфекцій. Продукція НК-клітин здійснюється кістковим мозком.

Ще одним механізмом первинного противірусного захисту є вироблення зараженими клітинами інтерферону [31]. Інтерферон діє на сусідні незаражені клітини і викликає у них несприятливість до зараження вірусами.

Основними механізмами специфічного набутої імунної відповіді є вироблення В-клітинами антитіл, що зв'язують вільні віруси, та знищення заражених клітин цитотоксичними Т-лімфоцитами.

При математичному запису описаних процесів використано закон діючих мас, згідно якому швидкості реакцій пропорційні добутку концентрації речовин, а також використано характеристики взаємодії та марківських процесів загибелі та відродження.

Дослідження фундаментальних і модельних результатів надає можливість на ґрунті моделей взаємодії антигенів та антитіл побудувати математичні моделі підтримки здоров'я людини. Основну роль у такому підході відіграє імунна система як комплекс програм з управління біонанооб'єктами в організмі людини.

Необхідність створення систем програмного забезпечення супроводу лікувального

процесу тісно пов'язана із розумінням стратегії імунної відповіді.

В організмі існує визначений рівень імунокомпетентних клітин (ИКК) зі своїми рецепторами -- антитілами, які є специфічні до всіх антигенів. Незначна кількість антигенів коли опиняється в організмі людини, починає розмножуватися у клітинах органу-мішені; частина антигенів скидається у кров, де взаємодіє із рецепторами ИКК -- із антитілами, утворюючи комплекси антигіло-антиген, які поглинаються антитіло-утворюючими клітинами. Через певний час після стимуляції в організмі формуються плазматичні клітини, які виробляють антитіла. Уражена частина органу не стимулює вихід плазматичних клітин, у результаті чого процес регенерації тканин органу-мішені запускається у відповідь на пошкодження. Механізм імунної відповіді зображено на рис. (1). На ньому позначено: ИКК+РА -- імунокомпетентні клітини із рецепторами-антитілами; КАА -- комплекси антиген-антитіло; ССОМ -- сигнал про стан органу- мішені (сигнал запуску програми формування плазматичних клітин); СРОМ -- сигнал про регенерацію органу- мішені (сигнал запуску програми регенерації органу-мішені); ПО-М - уражений орган-мішень; ПК -- плазматичні клітини.

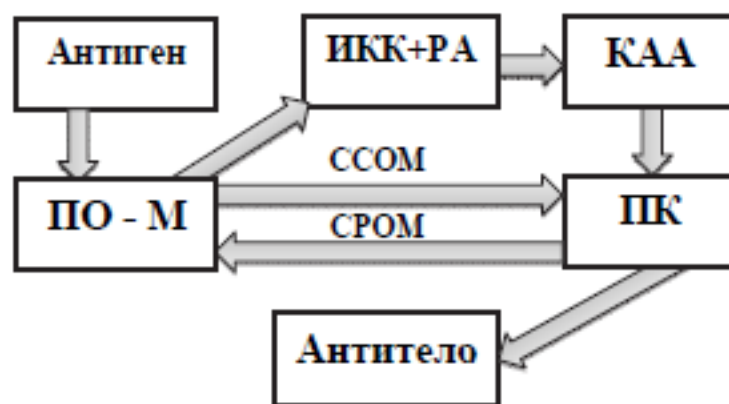


Рис. 7.1. Схема функціонування імунної відповіді

7.1.1 Математична модель протиінфекційного захисту

Розглянемо математичну модель функціонування імунної системи організму людини [4, 6,7]. Ця модель описує фундаментальні механізми імунного захисту, рис. 7.2..



Рис. 7.2. Простий механізм імунної реакції

Для формування базової моделі імунного захисту розглянемо такі положення імунології:

1. В організмі існує ненульовий рівень імунокомпетентних клітин зі своїми рецепторами -- антитілами, що є

- специфічні до всіляких антигенів, які вільно циркулюють як всередині клітин, так і за межами клітин;
2. Мала доза антигенів, опиняючись в організмі, починає розмножуватися у клітинах органу -мішені;
 3. Частина антигенів скидається у кров, де взаємодіє із рецепторами імунокомпетентних клітин -- із антитілами, утворюючи комплекси антитіло - антиген, які поглинаються клітинами, що утворюють антитіла і стимулюють імунну систему;
 4. Через час τ після стимуляції пропорційно до кількості комплексів антитіло- антиген в організмі формуються плазматичні клітини, які виробляють антитіла, специфічні до відповідних антигенів
 5. Зменшення числа плазматичних клітин та антитіл здійснюється також за рахунок їхнього старіння;
 6. Кількість плазматичних клітин, що утворюються у відповідь на антигенну стимуляцію, залежить від життєздатності ураженого органу;
 7. Процес регенерації тканин органу-мішені запускається у відповідь на пошкодження.

Перелічимо змінні моделі як неперервні функції:

- $V(t)$ -- концентрація антигенів, що розмножуються (патогенів) в ураженій частині органу-мішені [част./мл];
- $C(t)$ -- концентрація антиген-специфічних лімфоцитів (носіїв і продуцентів антитіл) у лімфоїдній тканині [част./мл];
- $F(t)$ -- концентрація антитіл (імуноглобулінів, рецепторів імунокомпетентних клітин тощо, що нейтралізують антигени) у крові, [кліт/мл];
- $m(t)$ -- частка зруйнованих антигеном клітин в ураженій частині органу-мішені, що впливає на послаблення життєздатності організму у ході захворювання, яке пов'язане із зменшенням активності органів, що забезпечують надходження імунологічного матеріалу: лейкоцитів, лімфоцитів, антитіл тощо, які необхідні для боротьби із антигенами, що розмножуються.

Математична модель будується на ґрунті балансу для кожної із залежних змінних за припущення, що "організм" описується однорідним замкнутим об'ємом, в якому усі компоненти процесу рівномірно змішані. Змінювання чисельності клітинних популяцій, що задіяні у процесі імунної відповіді розглядається на відріzkу часу $[t_0, T]$, де $t_0 = 0$ -- момент інфікування, враховується пошкодження і регенерація органів.

Табл.7.1. біологічний зміст параметрів

Параметр	Біологічний зміст
γ	Коефіцієнт, що враховує ймовірність зустрічі антигенів із антитілами
α	Коефіцієнт стимуляції імунної системи
	Стала швидкості
β	розмноження антигенів

μ_c	природного старіння плазматичних клітин
ρ	виробництва антитіл 1-ю плазматичною клітиною
η	розходу антитіл на знешкодження одиниці антигену
μ_f	природного руйнування антигену
σ	руйнування клітин органу-мішені антигеном
μ_m	регенерації органу-мішені
C^*	Передіснуючий рівень плазматичних клітин
τ	Час формування каскаду плазматичних клітин

Табл. 7.2. Значення параметрів моделі за різних форм захворювань

Параметр	Значення за різних форм захворювань				Розмірність
	субкл.	гостра	хрон.	летал.	
β	8	2	1	2	1/доба
γ	10	0,8	0,8	0,8	мл/(част доба)
α	10000	10000	1000	10000	кліт.мл/(част доба)
μ_c	0,5	0,5	0,5	0,5	1/доба
ρ	0,17	0,17	0,17	0,17	Мол/(кліт. Доба)
η	10	10	10	10	Мол/част
μ_f	0,17	0,17	0,17	0,5	1/доба
σ	10	10	10	700	мл/(част доба)
μ_m	0,12	0,12	0,12	0,12	1/доба
C^*	1	1	1	1	Кліт/мл
τ	0,5	0,5	0,5	0,5	доба

Рівень концентрації антигенів слугує основним показником динаміки хвороби і працездатності імунної системи. Маємо систему нелінійних диференціальних рівнянь із аргументом, що запізнюється, і розривною правою частиною, записані у нормальній формі Коші:

$$\begin{aligned}
 \frac{dV}{dt} &= \beta V - \gamma FV, \\
 \frac{dF}{dt} &= \rho C - \eta \gamma FV - \mu_f F, \\
 \frac{dC}{dt} &= \xi(m) \alpha F(t - \tau) V(t - \tau) - \mu_c (C - C^*), \\
 \frac{dm}{dt} &= \sigma V - \mu_m m,
 \end{aligned}
 \tag{20}$$

з початковими умовами

$$V(0) = V^0, \quad F(0) = F^0, \quad C(0) = C^0, \quad m(0) = m^0, \tag{21}$$

та фазовими обмеженнями:

$$V(t) \geq 0, \quad F(t) \geq 0, \quad C(t) \geq 0, \quad m(t) \geq 0, \tag{22}$$

де $\beta > 0$ -- швидкість (темп) розмноження антигенів; $\gamma > 0$ -- коефіцієнт, що враховує

ймовірність зустрічі вірусів із антитілами та силу їх взаємодії; $\alpha > 0$ -- коефіцієнт стимулювання імунної системи; $\rho > 0$ -- швидкість продукування антитіл однією плазмноклітиною; $\mu_c, \mu_f > 0$ -- величини, зворотні до тривалості життя плазмноклітин та антитіл відповідно; $\eta > 0$ -- кількість антитіл, що необхідне для нейтралізації одного вірусу; $\sigma > 0$ -- швидкість (темп) ураження органу; $\mu_m > 0$ -- швидкість відновлення маси ураженого органу; $C^* > 0$ -- попередній рівень імунокомпетентних клітин (плазмноклітин); $\tau > 0$ -- час, необхідний для формування каскаду плазмноклітин; $\xi(m)$ -- невід'ємна функція, що не зростає, яка враховує порушення нормальної роботи імунної системи внаслідок суттєвого ураження органу

$$\xi(m) = \begin{cases} 1, & 0 \leq m < 0,1 \\ (1-m) \cdot 10/9, & 0,1 \leq m \leq 1. \end{cases}$$

Модель (1)–(3) описує загальні закономірності, властиві усім інфекційним захворюванням. Фактично ця модель визначає вигляд залежності важкості інфекції від характеристик імунітету та властивостей антигену.

7.2 Керування імунологічними реакціями

Запровадимо у розгляд критерій якості J , що відображає імунний захист за рахунок мінімізації середньої швидкості пошкодження організму за фіксований час T :

$$J = \frac{1}{T} \int_0^T \sigma V(t) dt \rightarrow \min, \quad (23)$$

де $\sigma > 0$ -- швидкість (темп) ураження організму за рахунок антигенів.

Досягнення цієї мети пов'язане із проблемою управління імунологічними реакціями. В якості управління, що регулює інтенсивність процесів імунної відповіді і сприяє пришвидшеному відновленню уражених органів і тканин, задіюються при лікуванні засоби (зовнішнє управління), а також медіатори, що виробляються самим організмом (внутрішнє управління).

На ґрунті базової моделі інфекційного захворювання (1) побудуємо модель, що управляється, виходячи із таких припущень:

1. Управліннями є медичні препарати, що використовуються при лікуванні, тобто здійснюється зовнішнє управління шляхом введення додаткових членів у праві частини рівнянь моделі (1).

2. Визначений вид управління характерний для конкретної форми захворювання.

Для управління функціонуванням імунної системи при хронічних формах захворювання застосовується біостимуляція $u_1(t)$, що ґрунтується на введенні непатогенних антигенів. На рівні моделі, що визначає поведінку імунної системи, отримаємо модифікацію системи рівнянь (1) шляхом додавання трьох диференціальних рівнянь, що відображають характер процесу лікування:

$$\begin{aligned} \frac{dV_1}{dt} &= -\gamma_1 F_1 V_1 + u_1, \\ \frac{dF_1}{dt} &= \rho_1 C_1 - \eta_1 \gamma_1 F_1 V_1 - \mu_{f1} F_1, \\ \frac{dC_1}{dt} &= \frac{F_1 V_1}{FV + F_1 V_1} \alpha_1 F_1 (t - \tau) V_1(t - \tau) - \mu_{c1} (C_1 - C_1^*). \end{aligned} \quad (24)$$

Ефективність лікування гострих форм захворювання і захворювання із можливим летальним вибитком визначається за допомогою реалізації імунотерапії ($u_2(t)$), що ґрунтується на введенні готових імуноглобулінів або донорських антитіл, які надають можливість змінити динаміку хвороби на користь видоровлення. У результаті запровадження управляючої функції $u_2(t)$ модель, що визначає поведінку імунної системи, складається із рівнянь ((5)) із урахуванням модифікації динаміки функції $F(t)$, яка відображає характер лікування:

$$\frac{dF}{dt} = \rho C - \eta \gamma FV - \mu_f F + u_2 \quad (25)$$

Дослідження закономірностей протиінфекційного захисту виконується із залученням варіаційного принципу, що ґрунтується на припущенні, що поміж допустимих варіантів імунного захисту організму реалізуються ті, які мінімізують середню швидкість його пошкодження за фіксований час T .

$$I = \frac{1}{T} \int \sigma V(t) dt \rightarrow \min, \quad (26)$$

В якості керування, що регулює інтенсивність процесів імунної відповіді і сприяє прискореному відновленню уражених органів і тканин, виступають засоби, що використовують при лікуванні (зовнішнє управління), а також різні гормони і медіатори, що виробляються самим організмом (внутрішнє управління).

У результаті модель, що визначає поведінку імунної системи, складається із системи

$$\begin{aligned} \frac{dV}{dt} &= \beta V - \gamma FV, \\ \frac{dF}{dt} &= \rho C - \eta \gamma FV - \mu_f F + u_2, \\ \frac{dC}{dt} &= \begin{cases} G(t)\alpha F(t-\tau)V(t-\tau) - \mu_c(C-C^*), & 0 \leq m < 0,1 \\ \xi(m)G(t)\alpha F(t-\tau)V(t-\tau) - \mu_c(C-C^*), & 0,1 \leq m \leq 1 \end{cases}, \\ \frac{dm}{dt} &= \sigma V - \mu_m m, \\ \frac{dV_1}{dt} &= -\gamma_1 F_1 V_1 + u_1, \\ \frac{dF_1}{dt} &= \rho_1 C_1 - \eta_1 \gamma_1 F_1 V_1 - \mu_{f1} F_1, \\ \frac{dC_1}{dt} &= \begin{cases} G(t)\alpha_1 F_1(t-\tau)V_1(t-\tau) - \mu_{c1}(C_1-C_1^*), & 0 \leq m < 0,1 \\ \xi(m)G(t)\alpha_1 F_1(t-\tau)V_1(t-\tau) - \mu_{c1}(C_1-C_1^*), & 0,1 \leq m \leq 1, \end{cases} \end{aligned} \quad (27)$$

$$G(t) = \frac{F_1 V_1}{FV + F_1 V_1}.$$

Отже, оптимізаційна керована модель динаміки імунного захисту спрямована на

мінімізацію функціоналу (7) при динамічних умовах, заданих системою диференціальних рівнянь (8).

Початкові умови:

$$\begin{aligned} V(0) &= V^0, F(0) = F^0, C(0) = C^0, \\ V_1(0) &= V_1^0, F_1(0) = F_1^0, C_1(0) = C_1^0, m(0) = m^0. \end{aligned} \quad (28)$$

$$\begin{aligned} V(t) &\geq 0, F(t) \geq 0, C(t) \geq 0, \\ V_1(t) &\geq 0, F_1(t) \geq 0, C_1(t) \geq 0, m(t) \geq 0. \end{aligned} \quad (29)$$

Функції управління $u_1(t), u_2(t)$ характеризують швидкість введення біостимуляції і реалізації імунотерапії в одиницю часу відповідно і задовольняють обмеження:

$$0 \leq u_i(t) \leq b_i, \quad i = 1, 2, t \in [0, T], \quad (30)$$

де $b_i > 0$ -- максимальне число людей, що схильні до біостимуляції, імунотерапії в одиницю часу, що залежать від обмежених технічних і фінансових засобів.

Система рівнянь (8) є система із розривною правою частиною та з аргументом, що запізнюється, загальний вигляд якої

$$\frac{dx(t)}{dt} = \begin{cases} f_1(t, x(t), x(t-\tau), u(t)), & S(t, m) < 0, \\ f_2(t, x(t), x(t-\tau), u(t)), & S(t, m) \geq 0, \end{cases}$$

де $x = (V, F, C, m, V_1, F_1, C_1)$ -- абсолютно неперервна на $[0, T]$ вектор-функція стану; $u = (u_1, u_2)$ -- кусково-неперервна на $[0, T]$ -- кусково-неперервна на $[0, T]$ вектор-функція управління. Поверхня перемикання $S(t, m)$ -- неперервно диференційовна скалярна функція, що має вигляд $S(t, m) = m - m^*$, де $m^* = 0, 1$ -- граничний рівень ураження, за якого ще можлива нормальна робота імунної системи. Розглянемо випадок однократного "протикання" траєкторією поверхні переключення у точці τ_α (момент переключення), тобто точка, в якій $S(x(\tau_\alpha)) = 0$.

Для побудови оптимального управління використовується принцип максимуму.

Оскільки у задачі наявні фазові обмеження ((14)), забезпечимо їх виконання шляхом запровадження штрафного доданку у функціонал ((8)) і перейдемо до послідовності задач, в яких відсутні фазові обмеження, а функціонал має вигляд

$$I = \frac{1}{T} \int_0^T \sigma V(t) dt + A, \quad (31)$$

$$\begin{aligned} A = A_k \int_0^T & [(\max\{-V, 0\})^2 + (\max\{-C, 0\})^2 + (\max\{-F, 0\})^2 + (\max\{-m, 0\})^2 \\ & + (\max\{-V_1, 0\})^2 + (\max\{-C_1, 0\})^2 + (\max\{-F_1, 0\})^2] dt, \end{aligned}$$

$k = 1, 2, \dots, A_k > 0$ -- параметри штрафу.

Функція Понтрягіна для задачі (8), (9), (11), (12) має вигляд

$$H(t, x(t), x(t-\tau), \psi(t), \lambda_0) = \begin{cases} H_1(t, x(t), x(t-\tau), \psi(t), \lambda_0), & S(x) < 0, \\ H_2(t, x(t), x(t-\tau), \psi(t), \lambda_0), & S(x) > 0, \end{cases}$$

$$\begin{aligned} H_i(t, x(t), x(t-\tau), \psi(t), \lambda_0) = \\ -\lambda_0 \{ 1/T \sigma V(t) + A_k [(\max\{-V(t), 0\})^2 + (\max\{-C(t), 0\})^2 + (\max\{-F(t), 0\})^2 \\ + (\max\{-V_1(t), 0\})^2 + (\max\{-C_1(t), 0\})^2 + (\max\{-F_1(t), 0\})^2] \\ + (\psi_i(t), f_i(t, x(t), x(t-\tau), u(t))) \}, \end{aligned}$$

$$i = 1, 2; k = 1, 2 \dots A_k > 0, \lim_{k \rightarrow \infty} A_k = \infty.$$

Функція $\psi(t) = (\psi_1(t), \psi_2(t))$ -- сполучена вектор-функція, що визначена на $[0, \tau_\alpha] = T_1$ та $[\tau_\alpha, T] = T_2$ відповідно, неперервна і майже всюди неперервно диференційовна на цих відрізках.

Початкові умови:

$$V(0) = 10^{-6}, C(0) = 1, F(0) = 1, m(0) = 0, V_1(0) = 0, C_1(0) = 1;$$

$$F_1(0) = 1, u_1(0) = 10^{-3}, \gamma_1 = 0,5, \alpha_1 = 80,$$

$$\mu_{f_1} = 0,17, \rho_1 = 0,17, \mu_{c_1} = 0,5, \eta_1 = 20$$

при хронічній формі захворювання, та

$$V(0) = 10^{-6}, C(0) = 1, F(0) = 1, m(0) = 0, u_2(0) = 5$$

за гострої форми захворювання.

7.3 Математичне моделювання лікування інфекційних хворих вакцинацією

7.3.1 Огляд математичних моделей в імунології

Для аналізу найважливіших фізіологічних процесів при інфекційних захворюваннях широко використовується математичне моделювання, коли різні захворювання розглядаються з єдиних позицій як процес взаємодії імунної системи зі збудниками хвороби. Математичні моделі інфекційних захворювань являють собою нелінійні системи диференціальних рівнянь і містять значну кількість параметрів, що характеризують імунний статус організму і властивості антигену. В основу математичних моделей імунної відповіді покладено фундаментальні роботи Г. І. Марчука, у моделях якого відображено найсуттєвіші закономірності функціонування імунної системи при інфекційних захворюваннях.

Традиційні задачі у галузі математичного моделювання в імунології полягають у побудові та дослідженні моделей імунної відповіді та імунного захисту організму при інфекційних захворюваннях.

Інфекційне захворювання розглядається як конфлікт між імунною системою та популяцією збудників хвороби.

В якості фазових змінних моделі розглядаються такі найсуттєвіші характеристики захворювання:

1. Концентрація антигенів $V_f(t)$ у частині ураженого органу [част/мл].
2. Концентрація плазматичних клітин $C_v(t)$, [кліт/мл], -- це популяція продуцентів та носіїв антитіл (імунокомпетентні клітини).
3. Концентрація антитіл у крові $F(t)$, [част/мл]. Вони нейтралізують антигени.
4. Відносна характеристика ураженого органу $m(t)$, яку можна інтерпретувати як частку зруйнованих антигеном клітин організму. Це узагальнена характеристика пошкодження, яке антиген наносить органу-мішені, $m(t) = 1 - M(t)/M^*$, де $M(t)$ -- кількість клітин органу-мішені у момент часу t , V^* -- кількість клітин у нормі.

У табл. 7.3 наведено параметри нормального перебігу вірусного гепатиту Б.

Табл. 7.3. Параметри нормального перебігу ВГБ $M = \text{mol/l}$

Коеф.	Значення	Коеф.	Значення
	Концентрація		стимуляція
M^0	10^{-15} M	$b_H^{(E)}$	$4 \cdot 10^{16} \text{ 1/(M} \cdot \text{д.)}$
H_E^0	10^{-18} M	$b_H^{(B)}$	$4 \cdot 10^{16} \text{ 1/(M} \cdot \text{д.)}$
H_B^0	10^{-19} M	$b_p^{(E)}$	$5 \cdot 10^{33} \text{ 1/(M}^2 \cdot \text{д.)}$
E^0	10^{-18} M	$b_p^{(B)}$	$5 \cdot 10^{32} \text{ 1/(M}^2 \cdot \text{д.)}$
B^0	10^{-18} M	$b_p^{(P)}$	$5 \cdot 10^{32} \text{ 1/(M}^2 \cdot \text{д.)}$
P^0	$4,3 \cdot 10^{-22} \text{ M}$	nb_{CE}	$3,4 \cdot 10^{15} \text{ 1/(M} \cdot \text{д.)}$
h^0	$8,5 \cdot 10^{-14} \text{ M}$	b_{EC}	$1,6 \cdot 10^{14} \text{ 1/(M} \cdot \text{д.)}$
C^0	$0,5 \cdot 10^{-12} \text{ M}$	$b_p^{H_E}$	$5 \cdot 10^{30} \text{ 1/(M}^2 \cdot \text{д.)}$
V_F^0	$2,9 \cdot 10^{-16} \text{ M}$	$b_p^{H_B}$	$8 \cdot 10^{31} \text{ 1/(M}^2 \cdot \text{д.)}$
	швид. стимуляції		
α_M	$1,2 \text{ 1/day}$	γ_{MV}	$9,4 \cdot 10^9 \text{ 1/(M} \cdot \text{д.)}$
α_E	$0,4 \text{ 1/day}$	γ_{VM}	$1,6 \cdot 10^{15} \text{ 1/(M} \cdot \text{д.)}$
α_B	$0,1 \text{ 1/day}$	γ_{VF}	$3 \cdot 10^{11} \text{ 1/(M} \cdot \text{д.)}$
α_P	$0,4 \text{ 1/day}$	γ_{VC}	$4 \cdot 10^{15} \text{ 1/(M} \cdot \text{д.)}$
α_F	$0,043 \text{ 1/day}$	γ_{FV}	$8,6 \cdot 10^{11} \text{ 1/(M} \cdot \text{д.)}$
ρ_F	$1,7 \cdot 10^8$ мол/(д. кл.)	ρ_H^E	2
ρ_H^B	2	ρ_B	16
ρ_E	16	ρ_F	3
τ_H^E	0,6 day	τ_H^B	0,6 day
τ_E	2 day	τ_B	2 day
τ_P	3 day	b_m	0,05 1/day
n	5	α_m	0,121/day
σ	$2,5 \cdot 10^9$ 1/(M · д.)	ν	83 1/day

Імунна реакція проти впроваджених в організм збудників вірусних інфекцій (грип, корб, гепатит тощо) містить два типи імунної відповіді: гуморальної із виробленням антитіл В-системою лімфоцитів та клітинний із накопиченням цитотоксичних Т-лімфоцитів-ефекторів. Основний тип імунної відповіді, що забезпечує захист організму, є клітинний.

Цитотоксичні Т-лімфоцити-ефектори, що накопичилися внаслідок імунної відповіді по рецепторам на мембранах, виявляють заражені клітини і знищують їх, виконуючи роль кілерів клітин власного організму.

Математична модель використовується для аналізу та імітації імунологічних даних

і клініко-лабораторних даних стосовно вірусного гепатиту В.

Фізичний зміст змінних моделі

- $V_f(t)$ -- кількість повноцінних вірусних частинок (HBV) та 20 нм частинок HB_sAg в 1 мл обсягу кожного із блоків. Одиниці вимірювання концентрації: 1 част./мл або 1 Моль/мл = $6,02 \cdot 10^{23}$ част./мл. Вважається, що співвідношення концентрацій HBV/HB_sAg є стале і складає 10^2 -- 10^{-3} .

- $C_v(t)$ -- кількість заражених вірусами клітин в 1 мл печінки об'єму Q_{p2} .
- $m(t)$ -- кількість зруйнованих клітин в 1 мл печінки об'єму Q_{p2} .
- $P(t)$ -- кількість плазматичних клітин блоку ЛВ.
- $F(t)$ -- кількість молекул антитіл IgG в мл обсягу кожного із блоків.
- $M_v(t)$ -- кількість стимульованих макрофагів в 1 мл блоку ЛВ (лімфатичні вузли), що презентують антигени вірусу V у комплексі із SD -молекулами. Макрофаги стимулюють Ta -лімфоцити-ефектори E , які знищують клітини C_v клітин C , що заражені вірусом.

Передбачається, що кількість макрофагів M в організмі достатня для того, щоб M_v виникли у кількості, що пропорційна кількості вірусів.

- $H_E(t)$ -- кількість активованих T -лімфоцитів-хелперів, які виробляють інтерлейкіни-2 (ИЛ-2), що забезпечують проліферацію T -лімфоцитів-ефекторів в 1 мл блоку ЛВ (кл/мл або моль/мл).

- $H_B(t)$ -- кількість активованих T -лімфоцитів-хелперів, що забезпечують проліферацію B -лімфоцитів в 1 мл блоку ЛВ (кл/мл або моль/мл).

Стимуляція лімфоцитів-помічників (хелперів) H_B залежить від величини попереднього клону даної специфічності (тобто рецепторами проти комплексу Ia із даним антигеном), а також від кількості M_v .

- $E(t)$ -- кількість T -лімфоцитів-ефекторів в 1 мл блоку ЛВ (кл/мл або моль/мл).

Вважають, що клітина E під впливом двох сигналів -- від M_v та H_E -- дає початок клону клітин-ефекторів (E). що здатні до цитотоксичного (кілерного) впливу відносно C_v .

- $B(t)$ -- кількість B -лімфоцитів в 1 мл блоку.

7.3.2 Побудова моделі

При побудові рівнянь, що описують розвиток противірусної імунної відповіді клітинного типу, зроблено такі припущення.

1. Макрофаги M , що з'єднуються завдяки SD -комплексам із вірусами V , позначені як M_v , стимулюють T -лімфоцити-помічники, що позначаються як H_E , і T -лімфоцити-ефектори E . Цитотоксичні ефектори (E) знищують заражені вірусом клітини C_v , що є чутлива до даного вірусу тканини.

2. Вважається, що кількість макрофагів M в організмі досить велика для того, щоб M_v виникла у кількості, пропорційній числу вірусів.

3. Стимуляція T -помічників (H_E) залежить від величини попереднього клону даної специфічності (тобто із рецепторами проти комплексу SD із даним вірусом), а також від кількості M_V .

4. Вважається, що H_E діляться, тобто їхня кількість зростає. За достатньої кількості M_V , що з'явилися внаслідок ділення H_E , знову стимулюються.

5. Прийнято, що клітини E під впливом двох сигналів -- від M_V та H_E -- дають початок клону клітин-ефекторів (E), яким властиві цитотоксичні дії по відношенню до C_V .

6. Клітини чутливої до вірусу тканини несуть ідентичні SD -комплекси, які у клітинах, що заражені вірусом, стають SD .

Опис системи рівнянь імунної відповіді на вторгнення збудника вірусної природи.

Усі віруси у моделі вважаються "вільними", тобто такими, що вільно циркулюють поза клітинами у лімфі і плазмі крові, припускаючи, що усі внутріклітинні процеси описуються феноменологічно через коефіцієнт розмноження вірусів із виходом у плазму.

Відомо, що процеси активації, проліферації та диференцировки клітин імунної системи здійснюються у системі лімфатичних вузлів, що охоплюють орган-мішень, який вражається вірусом; процеси взаємодії вірусних частинок та антитіл -- у лімфі, крові та органу-мішені; процеси зараження, розмноження вірусів та руйнування заражених клітин органу локалізовані в органі-мішені.

7.3.3 Система рівнянь блоку органу-мішені

1. Запишемо рівняння балансу для числа вільних вірусів V_f (популяція вірусів в організмі). У цьому рівнянні відображено процеси синтезу та надходження у кров та лімфу вірусів і вірусних частинок, які зв'язують їх антитілами, зараження клітин органу-мішені і стимуляції макрофагів. Змінна $V_f(t)$ описує концентрацію 20 нм частинок HB_sAg у ЛВ, крові та печінки. Перерахунок кількості вірусів HBV гепатиту В при оцінці відповідних сталих здійснюється на основі співвідношення $1:(10^{-2} - 10^{-3})$ [79].

$$\frac{dV_f}{dt} = \nu C_V(t) + nb_{CE} C_V(t) E(t) - \gamma_{VM} V_f - \gamma_{VF} F V_f - \gamma_{VC} V_f(t) (C^* - C_V(t) - m(t)). \quad (32)$$

νC_V -- враховує збільшення популяції вірусів при їхньому розмноженні у заражених клітинах C_V . Самі вільні віруси стимулюють клітини M , переводять їх в M_V і потім знищуються антитілами. Коефіцієнт ν залежить від темпу відтворення вірусів із урахуванням внутріклітинних факторів захисту. У даній моделі передбачається, що нові віруси, які виникли у заражених клітинах, проникають у плазму і стають "вільними". Саме вільні віруси стимулюють клітини M , переводять їх у M_V і потім знищуються антитілами.

$nb_{CE} C_V E$ описує надходження у плазму вірусів, що вивільняються із заражених вірусами клітин C_V під час їхнього руйнування T -лімфоцитами-ефекторами E .

Третій і 4-й члени описують зменшення числа "вільних" вірусів за рахунок взаємодії із макрофагами M та антитілами F .

Останній член описує зменшення популяції V_f за рахунок впровадження вільних вірусів у здорові клітини і зараження їх, при цьому враховується скінченність клітин

органу-мішені, що досяжні для зараження.

Величини $\gamma_{V_M}, \gamma_{V_F}, \gamma_{V_C}$ вважаються сталими. Вони характеризують зворотні величини часів здійснення взаємодії вірусів V_f із макрофагами, антитілами і здоровими клітинами. Коефіцієнт ν характеризує кількість 20 нм частинок HB_sAg і вірусних частинок HBV за добу, що вивільняються у середньому при руйнуванні зараженої клітини T-лімфоцитом-ефектором. Швидкість синтезу HB_sAg за добу однією зараженою клітиною складає $10^{-15} \dots 10^{-13}$ г/доба, що відповідає $\nu \approx 10^3 \dots 10^5$ частинок HB_sAg та 1--10 вірусів HBV за добу.

Коефіцієнт n характеризує кількість 20 нм частинок HB_sAg та вірусних частинок HBV, що вивільняються T-лімфоцитом-ефектором. В якості оцінки n вибирають ту кількість HB_s -антигену, що виробляється за 5--7 діб, тобто $5 \cdot 10^{-15} \div 5 \cdot 10^{-12}$ г, що відповідає $10^3 \dots 10^6$ штук 20 нм частинок HB_sAg .

Коефіцієнт b_{CE} характеризує процес руйнування зараженої клітини T-ефектором.

Коефіцієнт γ_{VC} характеризує процес адсорбції вірусів незараженою клітиною органу-мішені.

2. Рівняння балансу заражених вірусом клітин $C_V(t)$:

$$\frac{dC_V}{dt} = \sigma V_f(t)(C^* - C_V - m(t)) - b_{CE} C_V(t) E(t) - b_m C_V(t). \quad (33)$$

Тут $C^* \approx 3 \cdot 10^8$ кл/мл.; $\sigma = 0,25 \cdot (10^{11} \div 10^{13})$ мл/(моль доба) -- стала швидкості зараження. b_{CE} -- стала швидкості руйнування заражених клітин T-ефектором; $b_{CE} \approx 6 \cdot 10^{15} \dots 4,3 \cdot 10^{20}$ мл/(моль доба); $b_m \approx (0,1 \div 0,005)$ 1/доба, характеризує процес руйнування заражених клітин за рахунок цитопатогенної дії вірусів.

3. Рівняння для частини органу m , що уражена вірусом:

$$\frac{dm}{dt} = b_{CE} C_V(t) E(t) + b_m C_V(t) - \alpha_m m(t), \quad \xi(m) = 1 - m/C^0. \quad (34)$$

Тут перший член враховує ефект елімінації ефекторами заражених вірусом клітин, другий -- незворотне ураження вірусами клітин C_V , які перестають функціонувати. Останній член описує відновлення уражених клітин за рахунок регенерації тканини (b_{CE}, b_m, α_m -- відповідні коефіцієнти).

Коефіцієнт $\alpha_m \approx 0,12$ 1/доба -- характеризує середню швидкість регенерації клітин печінки.

Модель ((18))--((206)) описує загальні закономірності, властиві усім інфекційним захворюванням.

Моделювання системи рівнянь органу-мішені.

Позначимо:

$$z_1(t) = \frac{V_f}{V_f^0}, z_2(t) = \frac{C_V}{C^0}, z_3(t) = \frac{m}{C^0}, z_4(t) = \frac{F}{F^0}, z_5(t) = \frac{P}{P^0}, \quad (35)$$

$$z_6(t) = \frac{E}{E^0}, z_7(t) = \frac{M_V}{M_V^0}, z_8(t) = \frac{H_E}{H_E^0}, z_9(t) = \frac{H_B}{H_B^0}, z_{10}(t) = \frac{B}{B^0}. \quad (36)$$

Запишемо ці рівняння у безрозмірній формі із урахуванням позначень (16), (17).

$$\frac{dz_1(t)}{dt} + a_{11}z_1(t) + a_{12}z_2(t) = N_{z_1}[y]; \quad (37)$$

$$N_{z_1}[y] = n_{1,1}z_2(t)z_6(t) + [n_{1,2}(z_2(t) + z_3(t)) + n_{13}z_4(t)]z_1(t); \quad (38)$$

$$\frac{dz_2(t)}{dt} + a_{22}z_2(t) + a_{21}z_1(t) = c_3 + N_{z_2}[y]; \quad (39)$$

$$N_{z_2}[y] = n_{2,1}(1 - z_3(t))z_1(t - \tau)[z_2(t - \tau) - z_3(t)] + n_{22}z_2(t - \tau)z_6(t); \quad (40)$$

$$\frac{dz_3(t)}{dt} + a_{33}z_3(t) + a_{32}z_2(t) = N_{z_3}[y]; \quad (41)$$

$$N_{z_3}[y] = n_{3,1}z_2(t)z_6(t); \quad (42)$$

$$\frac{dz_4(t)}{dt} + f_1z_4(t) + f_2z_5(t) = N_{z_4}[y]; \quad (43)$$

$$N_{z_4}[y] = n_{4,1}z_1(t)z_4(t) + n_{4,2}z_7(t)z_4(t); \quad (44)$$

$$\frac{dz_5(t)}{dt} + p_1z_5(t) = p_3 + N_{z_5}[y]; \quad (45)$$

$$N_{z_5}[y] = n_{5,1}(1 - z_3(t))z_7(t - \tau_p)z_9(t - \tau_p)z_{10}(t - \tau_p); \quad (46)$$

$$\frac{dz_6(t)}{dt} + e_1z_6(t) = e_3 + N_{z_6}[y]; \quad (47)$$

$$N_{z_6}[y] = n_{6,1}(1 - z_3(t))z_7(t - \tau_E)z_8(t - \tau_E)z_6(t - \tau_E) + n_{6,2}z_7(t)z_8(t)z_6(t) + n_{6,3}z_2(t)z_6(t). \quad (48)$$

Величини γ_{VM} , γ_{VF} , γ_{VC} характеризують зворотні величини часу взаємодії вільних вірусів із макрофагами, антитілами та здоровими клітинами відповідно.

Значення коефіцієнтів у цих рівняннях отримано із урахуванням рівнянь (20)–(24), позначень (16), (17) та табл. 7.3.

$$a_{11} = 1,14; a_{12} = -40; a_{13} = 1; a_{22} = b_m = 0,052, a_{21} = -0,0067; \quad (49)$$

$$a_{33} = 0,15, a_{32} = -b_m = -0,052; \quad (50)$$

$$f_1 = 0,4, f_2 = -0,86, f_3 = 0; r_1 = 0,4, r_2 = 0, r_3 = 0,4;$$

$$e_1 = 0,4, e_2 = 0, e_3 = 0,4; m_{v1} = a_{7,1} = 0,4;$$

Коефіцієнти при нелінійних складових:

$$n_{1,3} = 1,14, n_{1,1} = 0,2, n_{1,2} = -0,0731; n_{2,1} = 0,007, n_{2,2} = -0,007,$$

$$n_{2,3} = 0,0034; n_{5,1} = 1,19; n_{3,1} = 0,66 \cdot 10^{-3}, n_{4,1} = -0,25 \cdot 10^{-3},$$

$$n_{4,2} = -0,2142; n_{6,1} = 84,8, n_{6,2} = -5,3, n_{6,3} = -80; n_{7,1} = 0,0255;$$

$$n_{8,1} = 54, n_{8,2} = -27, n_{8,3} = -5,3 \cdot 10^{-6};$$

$$n_{9,1} = 54, n_{9,2} = -27, n_{9,3} = -0,8 \cdot 10^{-4};$$

$$n_{10,1} = 0,00272, n_{10,2} = -0,47 \cdot 10^{-3}.$$

Початкові умови

$$z_1(0) = 1, z_2 = 1, z_3(0) = 0, z_4(0) = 0,86, z_5(0) = 1, z_6(0) = 1,$$

$$z_7(0) = 1, z_8(0) = 1, z_9(0) = 1, z_{10}(0) = 1. \quad (51)$$

Рівень концентрації антигенів $V_f(t)$ слугує основним показником динаміки хвороби і працездатності імунної системи.

Оскільки у рівнянні ((213)) наявний вираз $b_{CE}C_v(t)E(t)$, який містить функцію $E(t)$, що визначається у блоці T -клітинної імунної відповіді, щоб зробити незалежними розв'язання рівнянь, які описують динаміку процесів у цих блоках, розглядатимемо тільки лінійну частину рівняння.

$$z_6^{(0)}(t) = \frac{1}{\alpha_E}(1 - e^{-\alpha_E t}) = 2,5(1 - e^{-0,4t}); \quad (52)$$

$$z_5^{(0)}(t) = \frac{1}{\alpha_F}(1 - e^{-\alpha_F t}) = 23,256(1 - e^{-0,043t}). \quad (53)$$

7.3.4 Вакцинація

Вплив вакцини на імунну відповідь

Вважається, що антитіла проти $HBsAg$ захищають від подальшої інфекції вірусом гепатиту В, і вони повинні бути викликані вакцинацією. Ін'єкції вакцини $HBsAg$ зазвичай вводять внутрішньо м'язово або підшкірно. Понад 95% вакцинованих (з трьома дозами, кожна з яких містить 40 мг $HBsAg$) у осіб розвивалися анти- HBs , які зберігалися впродовж усього спостереження за 24 місяці.

Щоб імітувати ефект вакцинації, до рівнянь (24)–(28) слід додати рівняння швидкості для концентрації, V_v , вакцини $HBsAg$: $z_7 = V_v/V_v^0$:

$$\frac{dz_7}{dt} + \gamma_{vM} z_7(t) = -\gamma_{vF} F^0 z_4 z_7; \quad (54)$$

Моделні рівняння для антигенпрезентуючих макрофагів M_v та anti- HBs імуноглобуліни F необхідно змінювати очевидним чином, щоб врахувати наявність антигену вакцини HBs : $z_8 = M_v = M_v/M_v^0$:

$$\frac{dz_8}{dt} + \alpha_M z_8(t) = \gamma_{MV} V_f^0 z_1 z_8 + \gamma_{MV} V_v^0 z_7 z_8; \quad (55)$$

$$\frac{dz_4}{dt} + \alpha_F z_4(t) = -\gamma_{FV} V_f^0 z_4 z_1 - \gamma_{FV} V_v^0 z_4 z_7. \quad (56)$$

$$V_f^0(t_0) = 10^3 \div 10^{10} \text{ част./мл.}$$

Кількісно конкретизуємо ситуацію, яку потрібно моделювати:

- (i) початкова доза вірусу гепатиту В, введена у кров, становить 50 HBV 1/л (V_f^0);
- (ii) частинки $HBsAg$ розміром $1,25 \cdot 10^{11}$ (част./мл) крові (V_v^0) вводяться одночасно під час вакцинації;
- (iii) початкова концентрація антитіл проти $HBsAg$ у крові відповідає концентрації до неімунного організму, тобто $F^* = 0,5 \text{ mIU/ml}$.

Час імунної відповіді такий, що виробляється значна кількість антитіл за 2 тижні після введення вакцини.

Зв'язок між дозою гепатиту В та інкубаційним періодом хвороби

Вірусний гепатит В має інкубаційний період, який у більшості випадків становить

45--120 днів. Фактори, які можуть бути причиною цих екстремальних показників захворювання вірусом гепатиту В включають:

- (i) розмір посівного матеріалу;
- (ii) шлях зараження;
- (iii) зміна вірусної патогенності хімічними або фізичними засобами;
- (iv) адміністрування специфічних антитіл; і
- (v) незвичайні взаємодії вірус-хазяїн.

Інкубаційний період для моделі оцінювався за час, необхідний для досягнення $V_f(t)$ свого максимального значення за 200 днів, тоді як тяжкість захворювання вважалася піковою величиною $m(t)$ за той самий інтервал часу. Відносні варіації $V_f(t_0)$ були у діапазоні $1 \div 10^{-7}$.

Розрахунки показали, що значення $m_{\max}$ не залежить від $V_f(t_0)$ у досліджуваних межах $V_f(t_0)$ у діапазоні $10^3 \div 10^{10}$ HBsAg част./мл і тому $1 \div 10^7$ HBV 1/мл.

Оцінка динамічного внеску окремих процесів у моделі гострого гепатиту

Метою наступного дослідження є отримання часової картини основних процесів, які починають діяти у різні проміжки часу під час гострого гепатиту В. Ідентифікація меж, як часових, так і кількісних, для процесів, які визначають, загалом, хід характеристик захворювання, надає додаткове уявлення про динамічний зв'язок між вірусологічними препаратами та імунологічними процесами.

Наприклад, шляхом додавання рівнянь для допоміжних змінних до системи моделі можна оцінити загальну кількість вірусних частинок, які секретують гепатоцити, частину антитіл, витрачених на нейтралізацію вірусних частинок, тощо. На жаль, експериментальні дані щодо цих характеристик дуже погані.

Ознайомимося з загальними характеристиками гострого гепатиту В і зв'яжемо їх з моделлю у такий спосіб:

$$V^C = \int_{t_0}^{t_0+T} \nu C_v dt,$$

загальна кількість HBsAg, що виділяється гепатоцитами, інфікованими вірусом, за час T

$$V^E = \int_{t_0}^{t_0+T} nb_{CE} C_v E dt,$$

загальна кількість HBsAg, що вивільняється з гепатоцитів, зруйнованих за час T цитотоксичними T-клітинами

$$V^M = \int_{t_0}^{t_0+T} \gamma_{VM} V_f M dt,$$

загальна кількість HBsAg, що поглинається організмами моноклеарною системою фагоцитів за час T

$$V^F = \int_{t_0}^{t_0+T} \gamma_{VF} V_f F dt,$$

загальна кількість HBsAg, нейтралізованого антитілами, специфічними до HBsAg за час T

$$F^P = \int_{t_0}^{t_0+T} \rho_F P dt,$$

загальна кількість антитіл, вироблених плазматичними клітинами за час T

$$F^V = \int_{t_0}^{t_0+T} \gamma_{FV} V_f F dt,$$

загальна кількість антитіл, зв'язаних з *HBsAg* за час T

$$C^V = \int_{t_0}^{t_0+T} \sigma V_f (C^* - C_V - m) dt,$$

загальна кількість інфікованих вірусом гепатоцитів за час T

$$C^E = \int_{t_0}^{t_0+T} n b_{CE} C_V E dt,$$

загальна кількість інфікованих гепатоцитів, зруйнованих цитотоксичними T-клітинами за час T

$$H_E^D = \int_{t_0}^{t_0+T} b_p^E [M_V(t) H_E(t) - H(t - \tau_H^E) M_V(t - \tau_H^E)] dt.$$

Ми оцінили ці характеристики, використовуючи розрахункові параметри, наведені в табл. 7.3 впродовж наступних інтервалів часу від початку зараження: $0 \leq t \leq 120$ днів, $110 \leq t \leq 120$ днів. Перший відповідає всьому періоду захворювання, тоді як друга характеризує фазу вираженої противірусної імунної відповіді. Аналізуючи дані табл. 7.1, можна зробити наступні висновки:

-- частка *HBsAg*-частинок, що вивільняються з гепатоцитів, інфікованих вірусом, руйнується цитотоксичними T-клітинами, є незначною і стає значущою лише в період вираженої імунної відповіді (110-120-й день після початку інфекції);

-- зважаючи на період захворювання (110-120 днів), можна легко помітити, що частинки *HBsAg* в основному нейтралізуються неспецифічними механізмами, але внесок гуморального імунітету, здається, стає помітним з наступним розвитком імунної відповіді;

-- провідний механізм звільнення організму від уражених вірусом гепатоцитів у перебігу гострого гепатиту В є T-клітинна опосередкована цитотоксичність, а не цитофатична дія вірусу гепатиту В.

-- враховуючи період захворювання від 110-120 днів, можна побачити, що частка антитіл, зв'язаних з частинками *HBsAg*, дуже мала, тоді як за інтервал 0-120 днів фракція антитіл, зв'язаних з *HBsAg*, розмір часток досягає 0,91;

-- порівняння кількості H^0 на 10-й і 120-й день після початку зараження, видно, що період імунної відповіді на розвиток гострого гепатиту В характеризується значною кількістю T-клітин, наявних у циклі поділу, при цьому 120-й день характеризується повним припиненням поділу T-клітин.

Тим часом такий аналіз на основі допоміжних характеристик моделі має виявити деякі проблеми: було виявлено, що одна цитотоксична T-клітина індукує в середньому знищення близько 100 клітин-мішеней.

Розв'язання системи рівнянь

Пошуку розв'язань систем диференціальних рівнянь із аргументом, що відхиляється, присвячена значна кількість наукових праць, наприклад, [2,9,13], у переважній більшості з яких застосовуються різницеві схеми типу Рунге--Кутта [11, 12, 14,19] відносно лінійної частини системи диференціальних рівнянь.

Метою роботи є розробка методу розв'язання систем нелінійних диференціальних рівнянь із аргументом, що відхиляється, що ґрунтується на ітераційному числово - аналітичному методі [5].

Запишемо цю систему рівнянь у векторній формі.

$$\frac{dy(t)}{dt} + Ay(t) = y_0(t) + N(y); \quad (57)$$

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & 0 \\ 0 & a_{3,2} & a_{3,3} & 0 & 0 \\ 0 & 0 & 0 & f_1 & f_2 \\ 0 & 0 & 0 & 0 & p_1 \end{bmatrix}; \quad (58)$$

$$y_0 = [0 \ a_{2,3} \ 0 \ 0 \ p_2]^T; \quad (59)$$

$$N = [N_{z_1} \ N_{z_2} \ N_{z_3} \ N_{z_4} \ N_{z_5}]^T; \quad (60)$$

Маємо задачу Коші для векторно-матричного рівняння (38) із початковими умовами (41)

Застосуємо до задачі (38), (41) інтегральне перетворення Лапласа.

Позначимо Y -- перетворення Лапласа шуканої функції. Тоді маємо

$$A(p)Y(p) = y_0/p + y(0) + \mathcal{L}[N(y)].$$

$$Y(p) = A^{-1}(p)[y_0/p + y(0)] + A^{-1}(p)\mathcal{L}[N(y)]. \quad (61)$$

$$A(p) = \begin{bmatrix} p + a_{1,1} & a_{1,2} & 0 & 0 & 0 \\ a_{2,1} & p + a_{2,2} & 0 & 0 & 0 \\ 0 & a_{3,2} & p + a_{3,3} & 0 & 0 \\ 0 & 0 & 0 & p + f_1 & f_2 \\ 0 & 0 & 0 & 0 & p + p_1 \end{bmatrix}. \quad (62)$$

Оскільки перетворення Лапласа є лінійне перетворення, пошук розв'язання цієї задачі здійснюємо за ітераційної процедурою. Позначимо $y^{(0)}(t)$ розв'язок лінійної частини задачі, тобто без урахування $N(y)$.

Цю систему рівнянь можна скоротити до трьох рівнянь, оскільки рівняння відносно $z_4(t)$ та $z_5(t)$ не залежать від перших трьох змінних. Тоді замість ((233)) матимемо

$$Y(p) = B^{-1}(p)[y_0/p + y(0)] + B^{-1}(p)\mathcal{L}[N(y)]. \quad (63)$$

$$B(p) = \begin{bmatrix} p + a_{1,1} & a_{1,2} & 0 \\ a_{2,1} & p + a_{2,2} & 0 \\ 0 & a_{3,2} & p + a_{3,3} \end{bmatrix}. \quad (64)$$

Маємо у лінійному наближенні:

$$Y^{(0)}(p) = B^{-1}(p)[y_0/p + y(0)], \quad (65)$$

Щоб отримати розв'язок у вигляді $z_q^{(0)}(t)$, треба знайти $B^{-1}(p)$. Маємо:

$$B^{-1}(p) = \frac{1}{\Delta_1(p)} \begin{bmatrix} (p+a_{2,2}) & -a_{1,2} & 0 \\ -a_{2,1} & p+a_{1,1} & 0 \\ \frac{a_{2,1}a_{3,2}}{p+a_{3,3}} & -\frac{a_{3,2}(p+a_{1,1})}{p+a_{3,3}} & \frac{\Delta_1(p)}{p+a_{3,3}} \end{bmatrix}. \quad (66)$$

$$\Delta_1(p) = p^2 + b_1 p + b_2, \quad b_1 = a_{1,1} + a_{2,2}, \quad b_2 = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

Запишемо у векторній формі початкові умови (41) із урахуванням збурень (40).

$$d(p) = [1 \quad b_m/p \quad 0]^T;$$

Маємо для лінійної частини векторного рівняння (42):

$$Y^{(0)}(p) = B^{-1}(p)d(p), \quad (67)$$

або у покомпонентному вигляді:

$$z_k^{(0)}(p) = \frac{h_{k,0}^{(0)}}{p} + \frac{h_{k,1}^{(0)}}{p+a_{3,3}} + \frac{h_{k,2}^{(0)} + h_{k,3}^{(0)}(p+\beta_k^{(0)})}{(p+\beta_k^{(0)})^2 + (\omega_k^{(0)})^2}. \quad (68)$$

$$\beta_k^{(0)} = b_1/2, \quad (\omega_k^{(0)})^2 = b_2 - (\beta_k^{(0)})^2.$$

$$z_k^{(0)}(t) = h_{k,0}^{(0)} + h_{k,1}^{(0)} e^{-a_{3,3}t} + g^{z_k}(t) \quad (69)$$

$$g^{z_k}(t) = e^{-\beta_k^{(0)}t} \begin{cases} [h_{k,2}^{(0)} \sin \omega_k^{(0)}t + h_{k,3}^{(0)} \cos \omega_k^{(0)}t], & (\omega_k^{(0)})^2 > 0, \\ [h_{k,2}^{(0)} \omega_k^{(0)}t + h_{k,3}^{(0)} \omega_k^{(0)}t], & (\omega_k^{(0)})^2 < 0, \end{cases} \quad (70)$$

$$h_{1,0}^{(0)} = -b_m/b_2 \cdot a_{1,2}, \quad h_{2,0}^{(0)} = b_m/b_2 \cdot a_{1,1}, \quad h_{1,1}^{(0)} = h_{2,1}^{(0)} = 0,$$

$$h_{1,3}^{(0)} = 1 - h_{1,0}^{(0)}, \quad h_{1,2}^{(0)} = a_{2,2} - b_1 h_{1,0}^{(0)} - h_{1,3}^{(0)} \beta_k^{(0)};$$

$$h_{2,3}^{(0)} = -h_{2,0}^{(0)}, \quad h_{2,2}^{(0)} = a_{2,2} - a_{2,1} - b_1 h_{2,0}^{(0)} - h_{2,3}^{(0)} \beta_k^{(0)};$$

$$h_{3,0}^{(0)} = -\frac{b_m a_{1,1}}{b_2 a_{3,3}}, \quad h_{3,1}^{(0)} = \frac{b_m - a_{2,1} + b_2 h_{3,0}^{(0)}}{a_{3,3}(b_1 - a_{3,3}) - b_2};$$

$$h_{3,3}^{(0)} = (a_{3,3} - b_1)h_{3,1}^{(0)} - b_1 h_{3,0}^{(0)}, \quad h_{3,2}^{(0)} = -(h_{3,0}^{(0)} + h_{3,1}^{(0)} + h_{3,3}^{(0)} \beta_k^{(0)})/\omega_k^{(0)}.$$

Далі використовуємо отриманий розв'язок у лінійному наближенні для визначення нелінійної частини $N(y)$ задачі.

Оскільки у нелінійні частини входять вирази для $E(t)$, $F(t)$ і неявно $P(t)$, запишемо розв'язки лінійних наближень для цих функцій (щоб забезпечити пошук автономного розв'язання системи рівнянь відносно органу-мішені).

$$z_6^{(0)}(t) = \frac{1}{\alpha_E} (1 - e^{-\alpha_E t}) = g^6(t) = h_{6,0}^{(0)} + h_{6,1}^{(0)}(\omega_6^{(0)}t) + h_{6,2}^{(0)}(\omega_6^{(0)}t); \quad (71)$$

$$h_{6,0}^{(0)} = 1/\alpha_E, \quad h_{6,1}^{(0)} = h_{6,2}^{(0)} = -h_{6,0}^{(0)};$$

$$[z_5^{(0)}(t) = \frac{r_3}{\alpha_P} (1 - e^{-\alpha_P t}) = g^5(t) = h_{5,0}^{(0)} + h_{5,1}^{(0)}(\omega_5^{(0)}t) + h_{5,2}^{(0)}(\omega_5^{(0)}t); \quad (72)$$

$$h_{5,0}^{(0)} = 1/\alpha_P, \quad h_{5,1}^{(0)} = h_{5,2}^{(0)} = -h_{5,0}^{(0)};$$

$$F^{(0)}(t) = g^4(t) = h_{4,0}^{(0)} + e^{-h_{4,4}^{(0)}t} [h_{4,1}^{(0)}(\omega_4^{(0)}t) + h_{4,2}^{(0)}(\omega_4^{(0)}t)] \quad (73)$$

$$g_0 = h_{4,0}^{(0)} = r_3/r_1, h_{4,2}^{(0)} = f_3 - g_0, h_{4,1}^{(0)} = (r_1 + f_1)g_0 + f_3r_1 - h_{4,2}^{(0)}h_{4,4}^{(0)},$$

$$h_{4,4}^{(0)} = \frac{r_1 + f_1}{2}, h_{4,3}^{(0)} = 1; (h_{4,5}^{(0)})^2 = r_1f_1 - (h_{4,4}^{(0)})^2 = -(r_1 - f_1)^2/4 \rightarrow \omega_4^{(0)} = \frac{r_1 - f_1}{2};$$

Тоді для рівняння (50) можна записати $k = \overline{1, 6}$, додавши останні три вирази.

Із урахуванням числових значень параметрів цієї системи побудуємо графіки отриманих розв'язків.

Перейдемо до пошуку розв'язків нелінійної системи рівнянь, що описують процеси в органі-мішені.

Із урахуванням отриманого розв'язку системи рівнянь у лінійному наближенні маємо такі вирази для нелінійних складових цієї системи:

$$N_{z_1} = n_{1,1}z_2^{(0)}(t) \cdot z_6^{(0)}(t) + z_1^{(0)}(t)[n_{1,2}(z_2^{(0)}(t) + z_3^{(0)}(t)) + n_{1,3}z_4^{(0)}(t)] = n_{1,1}[h_{2,0}^{(0)} + g^2(t)](1 - e^{-\alpha_E t}) + n_{1,2}[(h_{1,0}^{(0)} + g^1(t))(h_{2,0}^{(0)} + g^2(t)) + (h_{1,0}^{(0)} + g^1(t))(h_{3,0}^{(0)} + g^3(t))] + n_{1,3}(h_{1,0}^{(0)} + g^1(t))(h_{4,0}^{(0)} + g^4(t)). \quad (74)$$

Замістимо вирази для $g^1(t)$, $g^2(t)$, $g^3(t)$, $g^4(t)$ (розв'язок лінійної частини системи рівнянь) і виконаємо відповідні перетворення із використанням програми atbtp, виходом якої є вираз у просторі зображень за Лапласом.

$$\bar{N}(z_1) \approx \frac{z_2^1}{p} + \frac{z_0^1 p + z_1^1}{p^2 + z_4^1 p + z_3^1}.$$

Отримаємо:

$$N_{z_1}^{(0)}(t) = F(z_1^{(0)}, z_2^{(0)}, z_3^{(0)}, z_4^{(0)}) =$$

$$= r_0^1 + r_1^1 e^{-\alpha_m t} + r_6^1 e^{-2\beta_k^{(0)} t} + e^{-(\alpha_m + \beta_k^{(0)})t} [r_4^1 \cos \omega_k^{(0)} t + r_5^1 \sin \omega_k^{(0)} t]$$

$$+ e^{-\beta_k^{(0)} t} [r_2^1 \cos \omega_k^{(0)} t + r_3^1 \sin \omega_k^{(0)} t] + e^{-2\beta_k^{(0)} t} [r_7^1 \cos 2\omega_k^{(0)} t + r_8^1 \sin 2\omega_k^{(0)} t]. \quad (75)$$

Тепер до виразу (56) можна застосувати перетворення за Лапласом. Отримаємо:

$$\mathcal{L}[N_{z_1}] = \frac{r_0^1}{p} + \sum_{k=1}^4 \frac{v_{6k-6}^1 p + v_{6k-5}^1}{p^2 + v_{6k-2}^1 p + v_{6k-3}^1}. \quad (76)$$

Застосування до цього виразу алгоритму еквівалентного спрощення до ланцюга другого порядку призводить до такого виразу:

$$\mathcal{L}[N_{z_1}^{(0)}] \approx \frac{v_0^1}{p} + \frac{\bar{v}_0^1 p + \bar{v}_1^1}{p^2 + \bar{v}_4^1 p + \bar{v}_3^1} \quad (77)$$

де через $\bar{v}_k^2$, $k = \overline{0, 5}$ позначено коефіцієнти, що отримують у результаті спрощення дробово-раціонального виразу.

Нелінійна частина відносно $C_v(t)$.

Нелінійна частина відносно $C_v(t)$ має вигляд

$$N_{z_2}^{(0)} = \sigma z_1^{(0)}(t - \tau)[z_2^{(0)}(t - \tau) - z_3^{(0)}(t)] + n b_1^C z_2^{(0)}(t - \tau) E^{(0)}(t)$$

$$= u_0^2 + \sum_{k=1}^3 \{e^{-\beta_k^{(0)} t} [u_{3k-3}^2 + u_{3k-2}^2 \cos(\omega_k^{(0)} t - \tau^{(k)}) + u_{3k-1}^2 \sin(\omega_k^{(0)} t - \tau^{(k)})]\}; \quad (78)$$

Коефіцієнти u_j^2 , $j = \overline{1, 9}$, обчислюються за формулами, що отримано у результаті відповідних перетворень:

$$\beta_k^{(0)} = [\beta^{(0)}, 2\beta^{(0)}, 2\beta^{(0)}]; \quad \omega_k^{(0)} = [\omega^{(0)}, 2\omega^{(0)}, 2\omega^{(0)}], \quad \tau^{(k)} = [\tau, 2\tau, \tau];$$

$$r0 = a_0^2 + a_0^3; \quad \bar{u}_0^2 = a_0^1 r0; \quad u_{0,1}^2 = \frac{1}{2}(a_2^1 a_2^2 \pm a_3^1 a_3^2); \quad u_2^2 = \frac{1}{2}(a_2^1 a_3^2 + a_3^1 a_2^2);$$

$$u_{3,4}^2 = \frac{1}{2}(a_2^1 a_2^2 \pm a_3^1 a_3^2); \quad u_5^2 = \frac{1}{2}(a_2^1 a_3^2 + a_3^1 a_2^2);$$

$$u_6^2 = \frac{1}{2}[a_2^1(a_3^3 + a_2^3) + a_3^1(a_3^3 - a_2^3)]; \quad u_7^2 = \frac{1}{2}(a_2^1 a_2^3 - a_3^1 a_3^3); \quad u_8^2 = \frac{1}{2}(a_2^1 a_3^3 + a_3^1 a_2^3);$$

У просторі зображень цьому виразу відповідає

$$\bar{N}_{z_2}^{(0)}(p) = \frac{\bar{u}_0^2}{p} + \sum_{k=1}^3 e^{-\tau^{(k)} p} p \left[\frac{d_0^{(k)}}{p} + \frac{d_1^{(k)}(p + \beta_k^{(0)}) + d_2^{(k)} \omega_k^{(0)}}{(p + \beta_k^{(0)})^2 \pm (\omega_k^{(0)})^2} \right]. \quad (79)$$

Наявність множника затримки призводить до нелінійного дробово-раціонального виразу.

Апроксимуємо [5] множник $e^{-\tau p}$ також дробово-раціональним виразом, обмежуючись другим порядком апроксимації:

$$e^{-\tau^{(k)} p} \approx \frac{12 - 6\tau^{(k)} p + (\tau^{(k)})^2 p^2}{12 + 6\tau^{(k)} p + (\tau^{(k)})^2 p^2}. \quad (80)$$

$$\begin{aligned} \bar{N}_{z_2}(p) &= \frac{\bar{u}_0^2}{p} + \sum_{k=1}^3 \left[\frac{-12/\tau^{(k)} u_0^{(k)}}{p^2 + 6/\tau^{(k)} p + 12/(\tau^{(k)})^2} \right. \\ &\quad \left. + \frac{p^2 - 6/\tau^{(k)} p + 12/(\tau^{(k)})^2}{p^2 + 6/\tau^{(k)} p + 12/(\tau^{(k)})^2} \cdot \frac{u_1^{(k)} p + [u_1^{(k)} \beta_k^{(0)} + u_2^{(k)} \omega_k^{(0)}]}{p^2 + u_4^{(k)} p + u_3^{(k)}} \right]. \end{aligned} \quad (81)$$

$$\bar{u}_0^2 = u_0^{(1)} + u_0^{(2)} + u_0^{(3)}.$$

Або

$$\bar{N}_{z_2}(p) = \frac{\bar{u}_0^2}{p} + \sum_{k=1}^3 \frac{p^3 + r_1^{(k)} p^2 + r_2^{(k)} p + r_3^{(k)}}{p^4 + q_1^{(k)} p^3 + q_2^{(k)} p^2 + q_3^{(k)} p + q_4^{(k)}} \approx \frac{\bar{u}_0^2}{p} + \frac{\bar{v}_0^2 p + \bar{v}_1^2}{p^2 + \bar{v}_4^2 p + \bar{v}_3^2}. \quad (82)$$

У ((82)) запроваджено позначення:

$$\begin{aligned} r0 &= u_1^{(k)} \beta_k^{(0)} + u_2^{(k)} \omega_k^{(0)}, \quad q0 = 6/\tau^{(k)}; \quad q_0^{(k)} = 1, \quad q_1^{(k)} = q0 + u_4^{(k)}, \\ q_2^{(k)} &= q0(u_4^{(k)} + q0/3) + u_3^{(k)}, \quad q_3^{(k)} = q0(u_3^{(k)} + q0/3u_4^{(k)}), \quad q_4^{(k)} = q0^2/3u_3^{(k)}; \end{aligned} \quad (83)$$

$$\begin{aligned} r_0^{(k)} &= 1, \quad r_1^{(k)} = r0 - q0(u_1^{(k)} + 2u_0^{(k)}), \\ r_2^{(k)} &= q0(q0/3u_1^{(k)} - r0 - 2u_0^{(k)}u_4^{(k)}), \quad r_3^{(k)} = q0(r0q0/3 - 2u_0^{(k)}u_3^{(k)}); \end{aligned} \quad (84)$$

Отримаємо вираз для нелінійної частини рівняння відносно $m(t)$.

$$\begin{aligned} N_{z_3} &= N_m = b_1^C C_v(t) E(t) = b_1^C C_v^{(0)}(t) E^{(0)}(t) \\ &= b_1^C [h_{3,0}^{(0)} + h_{3,1}^{(0)} e^{-\alpha_m t} + g_3(t)] E^0 (1 - e^{-\alpha_E t}) \\ &= b_1^C E^0 [h_{3,0}^{(0)} - h_{3,0}^{(0)} e^{-\alpha_E t} + h_{3,1}^{(0)} e^{-\alpha_m t} + h_{3,1}^{(0)} e^{-(\alpha_m + \alpha_E)t} + g_3(t) - e^{-\alpha_E t} g_3(t)] \end{aligned} \quad (85)$$

У просторі зображень цьому виразу відповідає:

$$\mathcal{L}[N_{z_3}(t)] = \frac{\bar{v}_0^3}{p} + \sum_{k=0}^2 \frac{v_{6k}^3 p + v_{6k+1}^3}{p^2 + v_{6k+4}^3 p + v_{6k+3}^3} \approx \frac{\bar{v}_0^3}{p} + \frac{\bar{v}_0^3 p + \bar{v}_1^3}{p^2 + \bar{v}_4^3 p + \bar{v}_3^3}. \quad (86)$$

Коефіцієнти v_j^3 , $j = \overline{1, 18}$, обчислюються так:

$$\begin{aligned}\bar{v}1_0^3 &= h_{3,0}^{(0)}, v_0^3 = 2, v_1^3 = 2\alpha_m + \alpha_E, v_4^3 = 2\alpha_m + \alpha_E, v_3^3 = \alpha_m(\alpha_m + \alpha_E), \\ v_6^3 &= v_{12}^3 = h_{3,2}^{(0)}, v_7^3 = h_{3,2}^{(0)}\beta_k^{(0)} + h_{3,3}^{(0)}\omega_k^{(0)}, v_{10}^3 = 2\beta_k^{(0)}, v_9^3 = (\omega_k^{(0)})^2 + (\beta_k^{(0)})^2, \\ v_{13}^3 &= h_{3,2}^{(0)}(\beta_k^{(0)} + \alpha_E) + h_{3,3}^{(0)}\omega_k^{(0)}, v_{15}^3 = (\omega_k^{(0)})^2 + (\beta_k^{(0)} + \alpha_E)^2, v_{16}^3 = 2(\beta_k^{(0)} + \alpha_E).\end{aligned}$$

$$\mathcal{L}[N_{z_4}(t) = N_F(t)] = \mathcal{L}_t[n_{4,1}z_4^{(0)}z_1^{(1)} + n_{4,2}z_4^{(0)}z_7^{(0)}] = \frac{r_2^4}{p} + \frac{r_0^4 + r_1^4 p}{p^2 + r_4^4 p + r_3^4}. \quad (87)$$

$$\mathcal{L}[N_{z_7}(t) = M_V(t)] = \mathcal{L}_t[n_{7,1}z_4^{(1)}z_7^{(0)}] = \frac{r_2^7}{p} + \frac{r_0^7 + r_1^7 p}{p^2 + r_4^7 p + r_3^7}. \quad (88)$$

Коефіцієнти r_j^k обчислюються у програмі, що наведена нижче.

Замість отриманих виразів у (49).

$$Y^{(m)}(p) = Y^{(0)}(p) + B^{-1}(p)N_y^{(m-1)}(p) = Y^{(0)}(p) + V(p), \quad m = 0, 1, 2, \dots \quad (89)$$

$$V(p) = \begin{bmatrix} \frac{p+b_m}{\Delta_1} & \frac{v}{\Delta_1} & 0 \\ \frac{\alpha C^0}{\Delta_1} & \frac{p+a_{V_f}}{\Delta_1} & 0 \\ \frac{\alpha C^0 b_m}{(p+\alpha_m)\Delta_1} & \frac{b_m(p+a_{V_f})}{(p+\alpha_m)\Delta_1} & \frac{1}{p+\alpha_m} \end{bmatrix} \cdot \begin{bmatrix} \frac{r_2^1}{p} + \frac{r_1^1 p + r_0^1}{p^2 + r_4^1 p + r_3^1} \\ \frac{r_2^2}{p} + \frac{r_1^2 p + r_0^2}{p^2 + r_4^2 p + r_3^2} \\ \frac{r_2^3}{p} + \frac{r_1^3 p + r_0^3}{p^2 + r_4^3 p + r_3^3} \end{bmatrix}. \quad (90)$$

$$\Delta_1 = p^2 + b_1 p + b_2.$$

Після виконання відповідних дій у цьому виразі, переходу у простір оригіналів і додавання лінійних частин згідно з (70) отримаємо розв'язок системи рівнянь у першому наближенні у такому вигляді.

$$\begin{aligned}z_k^{(1)}(t) &= q1^k + q2^k e^{\alpha_m t} + e^{-\beta_k^{(0)} t} [q_1^{(k)} \cos \omega_k^{(0)} t + q_2^{(k)} \sin \omega_k^{(0)} t] \\ &+ e^{-\gamma_k t} [v_1^{(k)} \cos \omega_k^{(0)} t + v_2^{(k)} \sin \omega_k^{(0)} t], \quad k = \overline{1, 7}.\end{aligned} \quad (91)$$

У цих виразах враховано розв'язок системи рівнянь у лінійному наближенні. Подальші обчислення виконуються за аналогічними алгоритмами, що містять подання добутків дробово-раціональних виразів у просторі зображень за Лапасом у вигляді доданків найпростіших ланцюгів (першого та другого степеню), приведення подібних і запис результатів у просторі оригіналів у вигляді, аналогічному (72).

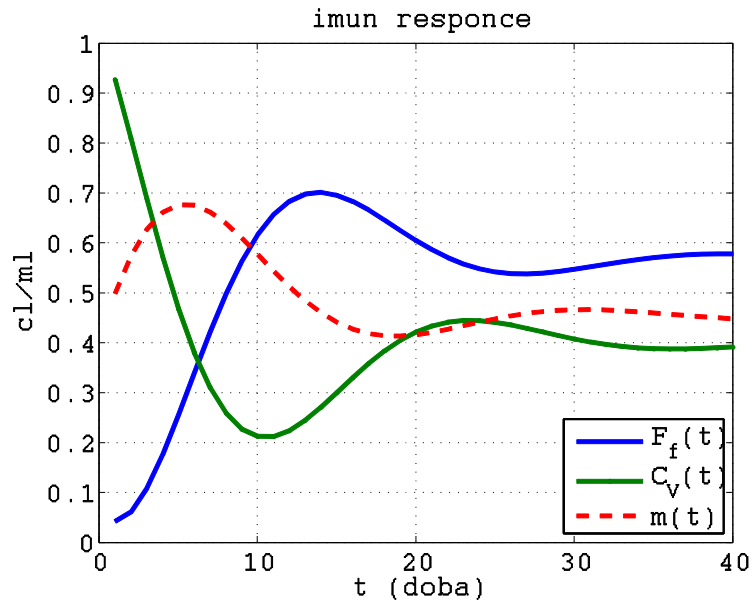


Рис. 7.3. Результати моделювання органу--мішені

7.4 Регуляція протівірусної відповіді

7.4.1 Математична модель

1. Здорові клітини C_Z .

Рівняння, що визначає швидкість змінювання кількості здорових клітин при вірусній інфекції та імунної відповіді:

$$\frac{dC_Z}{dt} = -k_1(C_Z + C_R)C_I + k_2C_R - k_3C_ZC_{IFN} - k_4C_ZC_V, \quad (92)$$

де C_Z -- кількість здорових клітин органу-мішені; C_R -- кількість стійких (резистентних) клітин органу-мішені; C_D -- кількість мертвих клітин органу-мішені; C_{IFN} -- концентрація інтерферону; C_V -- концентрація вірусів.

Коефіцієнт відновлення (регенерації) клітин епітелію респіраторного тракту $k_1 = 4$ [1/доба]; $k_2 = 1$ [1/доба]; $k_3 = 0,01$ [1/доба]; $k_4 = 0,34$ [1/доба].

Середня тривалість життя клітин епітелію респіраторного тракту складає 14 діб.

2. Інфіковані клітини C_I .

$$\begin{aligned} \frac{dC_I}{dt} + k_7C_I = & k_4C_ZC_V - k_5C_{NK}C_IC_{IL2}(1 - k_8F(C_K)) \\ & - k_6C_{CTL}C_I(1 - k_9F(C_K))1(C_{CNL} - k_{46}), \end{aligned} \quad (93)$$

де C_{NK} -- концентрація НК-клітин (природні кіллери); C_I -- кількість інфікованих клітин органу-мішені; C_{IL2} -- концентрація інтерлейкіну-2; C_K -- концентрація кортизолу; C_{CTL} -- концентрація цитотоксичних Т-лімфоцитів.

$$F(C_K) = \frac{C_K(t-T)}{1 + C_K(t-T)}.$$

Член $k_5C_{NK}C_IC_{IL2}(1 - k_8F(C_K))$ описує руйнування інфікованих клітин органу-

мішені за допомогою НК-клітин без утворення нових вірусів. Руйнування здійснюється під час зустрічі інфікованої клітини з НК-клітиною. Інтерлейкін-2 стимулює вплив на активність НК-клітин. Кортизол пригнічує активність НК-клітин, вплив кортизолу має характер запізнення, оскільки необхідний час для його доставки із нирок до органу-мішені. $T = 0,0132$ [доба]. Коефіцієнт $k_5 = 0,05$ [1/доба] відображає руйнування інфікованих клітин НК-клітинами, що стимульовані інтерлейкіном-2. Ступінь пригнічення кортизолом НК-клітин визначається коефіцієнтом $k_8 = 0,5$.

Член $k_6 C_{CTL} C_I (1 - k_9 F(C_K)) 1(C_{CNL} - k_{46})$ описує процес руйнування інфікованих клітин за допомогою цитотоксичних Т-лімфоцитів. Для опису ефекту накопичення даного типу клітин у лімфатичних вузлах до досягнення визначеного рівня використовується функція Хевісайда $1(t)$, $k_{46} = 1$. Коефіцієнт $k_6 = 0,066(0,02 \div 0,1)$ [1/доба] відображає швидкість руйнування інфікованих клітин цитотоксичними Т-лімфоцитами. Чим вище значення коефіцієнту, тим коротша тривалість захворювання і пошкодження органу-мішені.

Ступінь пригнічення кортизолом цитотоксичних Т-лімфоцитів визначається коефіцієнтом $k_9 = 0,5$. Коефіцієнт $k_7 = 1,5$ [1/доба] визначає швидкість переходу інфікованої клітини у мертвий стан.

Стійкі клітини C_R .

$$\frac{dC_R}{dt} + k_2 C_R = k_3 C_Z C_{IFN}, \quad (94)$$

Перехід здорових клітин у стійкий стан внаслідок впливу на них інтерферону описується членом $k_3 C_Z C_{IFN}$.

Інтерферон C_{IFN} .

$$\frac{dC_{IFN}}{dt} + k_{12} C_{IFN} = k_{10} C_I - k_{11} C_Z C_{IFN} \quad (95)$$

Інфіковані клітини стимулюють уроджений імунітет за рахунок виділення молекул інтерферону α і β , що взаємодіють зі здоровими клітинами, які набувають властивість стійкості до інфекції. Цей механізм надає організму необхідний час для вироблення адаптивної імунної відповіді та повної ліквідації вірусу. $k_{10} = 2000$ [1/доба], $k_{11} = 17$ [1/доба], $k_{12} = 8$ [1/доба].

Мертві клітини C_D

$$\begin{aligned} \frac{dC_D}{dt} = & k_7 C_I + k_5 C_{NK} C_{IL2} (1 - k_8 F(C_K)) + k_6 C_{CTL} C_I (1 - k_9 F(C_K)) \\ & \times 1(C_{CTL} - k_{46}) - k_{13} C_D C_M, \end{aligned} \quad (96)$$

де C_M -- концентрація макрофагів (моноцитів).

Час затримки пригнічуючої дії кортизолу на активність НК-клітин складає 19 хвилин ($T = 0,0132$ [доба]. $k_5 = 0,05$ [1/доба]. Ступінь пригнічення кортизолом НК-клітин задається коефіцієнтом $k_8 = 0,5$.

Опис процесу руйнування інфікованих клітин за допомогою цитотоксичних Т-лімфоцитів задається членом $k_6 C_{CTL} C_I (1 - k_9 F(C_K)) 1(C_{CTL} - k_{46})$. Для опису ефекту затримки даного типу клітин у лімфатичних вузлах до досягнення визначеної кількості використовується функція Хевісайда $1(C_{CTL} - k_{46})$, $k_{46} = 1$. $k_6 = 0,066$, $k_9 = 0,5$, коефіцієнт

$k_{13} = 10$ [1/доба] характеризує знищення мертвих клітин макрофагами.

Антиген C_V

$$\frac{dC_V}{dt} + k_{17}C_V = k_{14}C_I - k_{15}C_VC_A - k_{16}C_VC_Z, \quad (97)$$

де C_A -- концентрація антитіл.

Вироблення вірусів інфікованими клітинами описується членом $k_{14}C_I$, $k_{14} = 510$ [1/доба] -- коефіцієнт виділення вірусів інфікованими клітинами органу-мішені. Значення k_{14} визначається із того факту, що одна інфікована клітина виділяє порядку $10^3 \div 10^4$ вірусів за добу.

$k_{15}C_VC_A$ описує процес нейтралізації вірусів специфічними антитілами.

Швидкість взаємодії вірусів та антитіл -- $k_{15} = 619,2$ [1/доба]. Значення коефіцієнту ґрунтується на тому факті, що для нейтралізації одного вірусу необхідно 10 антитіл. $k_{16} = 1,02$ [1/доба], $k_{17} = 1,7$ [1/доба].

Макрофаги C_M

$$\frac{dC_M}{dt} + k_{19}C_M = k_{18}F_b, \quad (98)$$

F_b -- функціональна здатність кісткового мозку, синтезуюча функція.

Коефіцієнт вироблення макрофагів із урахуванням зниження інтенсивності кровотворної функції кісткового мозку $k_{18} = 0,03$ [1/доба]. $k_{19} = 0,03$ [1/доба].

Життєвий цикл макрофагів складає від 14 до 60 діб, після чого вони руйнуються і видаляються печінкою.

Інтерлейкін-1 C_{IL1}

$$\frac{dC_{IL1}}{dt} + k_{22}C_{IL1} = k_{20}C_MC_D(1 - k_{21}F(C_K)), \quad (99)$$

Під час взаємодії із мертвими клітинами макрофаги виділяють у кров інтерлейкін-1. Коефіцієнт $k_{20} = 0,145$ [1/доба]. Ступінь пригнічення кортизолом функції макрофагів визначається коефіцієнтом $k_{21} = 0,5$. Коефіцієнт природного розпаду інтерлейкіну-1 -- $k_{22} = 0,1245$ [1/доба].

Т-хелпери C_{TH}

$$\frac{dC_{TH}}{dt} + k_{24}C_{TH} = k_{23}F_b, \quad (100)$$

Коефіцієнт $k_{23} = 0,0058$ [1/доба] визначає інтенсивність вироблення кістковим мозком Т-хелперів. Життєвий цикл макрофагів складає близько 170 діб, після чого вони руйнуються. Коефіцієнт $k_{24} = 0,0058$ [1/доба] визначає природне виведення Т-хелперів.

НК-клітини C_NK

$$\frac{dC_{NK}}{dt} + k_{28}C_{NK} = k_{27}F_b, \quad (101)$$

НК-клітини задіяні у захисті організму проти вірусних збудників патогенів. Вони знищують усі клітини, які не проходять розпізнавання за принципом "свій-чужий". Коефіцієнт $k_{27} = 0,11$ [1/доба] визначає вироблення кістковим мозком НК-клітин. Коефіцієнт $k_{28} = 0,11$ [1/доба] характеризує природну загибель НК-клітин.

Інтерлейкін-2 C_{IL2}

Запуск активного ділення клітин імунної відповіді здійснюється після перших сигналів організму про наявність вірусної інфекції через стимулюючий вплив інтерлейкіну-2.

$$\frac{dC_{IL2}}{dt} + k_{26}C_{IL2} = k_{25}C_{TN}C_{IL1}, \quad (102)$$

$K_{25} = 0,328$ [1/доба] характеризує вироблення Т-хелперами інтерлейкіну-2. Коефіцієнт $k_{26} = 0,248$ [1/доба] визначає швидкість природного розпаду інтерлейкіну-2.

Цитотоксичні Т-лімфоцити C_{CTL}

$$\frac{dC_{CTL}}{dt} + k_{32}C_{CTL} = k_{29} + k_{30}C_{CTL}C_{IL2} - k_{31}C_{CTL}C_I(1 - k_9F(C_K))1(C_{CTL} - k_{46}). \quad (103)$$

Коефіцієнт $k_{29} = 0,4$ [1/доба] характеризує механізм природного (фонового) вироблення цитотоксичних Т-лімфоцитів, що необхідний для підтримки мінімального рівня клітин за відсутності інфекції. Коефіцієнт $k_{30} = 8,3$ [1/доба] ділення цитотоксичних Т-лімфоцитів за наявності інтерлейкіну-2. Коефіцієнт $k_{31} = 2,72$ [1/доба] відображає руйнування цитотоксичних Т-лімфоцитів при взаємодії із інфікованими клітинами. Коефіцієнт $k_{32} = 0,4$ [1/доба] характеризує природу загибель цитотоксичних Т-лімфоцитів.

Рівняння, що описує змінювання концентрації В-клітин, можна записати як

$$\frac{dC_B}{dt} + k_{34}C_B = k_{46} + k_{33}C_B C_{IL2}. \quad (104)$$

Коефіцієнт ділення В-клітин за наявності інтерлейкіну-2 приймається рівним $k_{33} = 11,5$ [1/доба]. Коефіцієнт $k_{34} = 0,4$ [1/доба] характеризує природну загибель В-клітин.

Антитіла.

Рівняння, що характеризує змінювання концентрації антитіл, можна подати у вигляді

$$\frac{dC_A}{dt} + k_{38}C_A = k_{35}C_B(1 - k_{36}F(C_K))1(C_B - k_{45}) - k_{37}C_V C_A. \quad (105)$$

Антитіла відіграють ключову роль у боротьбі із вірусом. Вони нейтралізують його до того, як він інфікує клітини органу-мішені, а НК-клітини і цитотоксичні Т-лімфоцити відповідають за видалення інфікованих клітин і запобігають розмноженню вірусу.

$k_{35}C_B(1 - k_{36}F(C_K))1(C_B - k_{45})$ характеризує вироблення антитіл В-клітинами. Коефіцієнт $k_{35} = 0,043$ [1/доба], ступінь пригнічення кортизолом В-клітин задається коефіцієнтом $k_{36} = 0,5$. Нейтралізація вірусів антитілами визначається коефіцієнтом $k_{37} = 0,043$ [1/доба]. Коефіцієнт природного розпаду антитіл дорівнює $k_{38} = 0,043$ [1/доба].

Кортизол.

$$\frac{dC_K}{dt} + k_{44}C_K = k_{49}F_a C_{ACTH}(t - T). \quad (106)$$

Коефіцієнт $k_{44} = 0,1972$ [1/доба] природного виведення адренокортикотропного гормону.

Система рівнянь імунної відповіді:

$$\frac{dC_Z}{dt} = -k_1(C_Z + C_R)C_I + C_Z[k_2C_R - k_3C_{\text{IFN}} - k_4C_V], \quad (107)$$

$$\frac{dC_I}{dt} + k_7C_I = k_4C_ZC_V - C_Ik_6C_{\text{CTL}}(1 - k_9F(C_K))1(C_{\text{CTL}} - k_{46}), \quad (108)$$

$$\frac{dC_R}{dt} + k_2C_R = k_3C_ZC_{\text{IFN}}, \quad (109)$$

$$\frac{dC_{\text{IFN}}}{dt} + k_{12}C_{\text{IFN}} = k_{10}C_I - k_{11}C_ZC_{\text{IFN}} \quad (110)$$

$$\frac{dC_V}{dt} + k_{17}C_V = k_{14}C_I - C_V[k_{15}C_A - k_{16}C_Z], \quad (111)$$

$$\frac{dC_{\text{CTL}}}{dt} + k_{32}C_{\text{CTL}} = k_{29} + C_{\text{CTL}}[k_{30}C_{\text{IL2}} - k_{31}C_I(1 - k_9F(C_K))1(C_{\text{CTL}} - k_{46})]. \quad (112)$$

$$\frac{dC_A}{dt} + k_{38}C_A = k_{35}C_B(1 - k_{36}F(C_K))1(C_B - k_{45}) - k_{37}C_VC_A;$$

$$\frac{dC_A}{dt} + k_{38}C_A = -k_{37}C_VC_A. \quad (113)$$

$$F(C_K) = \frac{C_K(t-T)}{1 + C_K(t-T)}.$$

Рівняння нейроендокринної системи при вірусній інфекції:

$$\frac{dC_B}{dt} + k_{34}C_B = k_{46} + k_{33}C_BC_{\text{IL2}}. \quad (114)$$

$$\frac{dC_{\text{IL1}}}{dt} + k_{22}C_{\text{IL1}} = -k_{20}C_MC_I(1 - k_{21}F(C_K)), \quad (115)$$

$$\frac{dC_{\text{IL2}}}{dt} + k_{26}C_{\text{IL2}} = k_{25}C_{\text{TN}}C_{\text{IL1}}, \quad (116)$$

$$\frac{dC_M}{dt} + k_{19}C_M = k_{18}F_b, \quad (117)$$

$$\frac{dC_{\text{TH}}}{dt} + k_{24}C_{\text{TH}} = k_{23}F_b, \quad (118)$$

$$\frac{dC_{\text{NK}}}{dt} + k_{28}C_{\text{NK}} = k_{27}F_b, \quad (119)$$

$$\frac{dC_K}{dt} + k_{44}C_K = k_{49}F_aC_{\text{ACTH}}(t-T). \quad (120)$$

Система рівнянь доповнюється початковими умовами:

$$\begin{aligned} C_Z(0) = C_Z^0, C_I(-T, 0) = C_I^0, C_V(0) = C_V^0, C_{\text{IFN}}(0) = C_{\text{IFN}}^0, C_D(-T, 0) = C_D^0, \\ C_M(0) = C_M^0, C_{\text{IL1}}(0) = C_{\text{IL1}}^0, C_{\text{IL2}}(0) = C_{\text{IL2}}^0, C_{\text{CTL}}(-T, 0) = C_{\text{CTL}}^0, \\ C_B(0) = C_B^0, C_A(-T, 0) = C_A^0, C_K(-T, 0) = C_K^0. \end{aligned} \quad (121)$$

7.5 Моделі динаміки гепатиту С

7.5.1 Вступ

Вірус гепатиту С (HCV, ВГС) вражає лише клітини печінки людини (і деяких вищих приматів). Незважаючи на те, що клітини печінки, інфіковані HCV (гепатоцити), спочатку можуть залишатися життєздатними, кульмінаційний стрес для печінки через 10–20-річний період інфікування призводить до розвитку цирозу печінки та раку печінки. Ураження печінки, викликане HCV-інфекцією, є основною причиною для трансплантації печінки в західному світі. Значні зусилля, щоб зрозуміти механізми реплікації ВГС [1–3], досягли кульмінації у розробці ефективних противірусних препаратів прямої дії (DAA) [4] проти HCV, які зараз пропонують значні показники лікування для багатьох пацієнтів, інфікованих вірусом.

Але багато аспектів реплікації ВГС залишаються недостатньо вивченими. Зокрема, хоча різні центральні етапи реплікації геному вірусу можуть бути виявлені експериментально, ще багато процесів розуміються лише частково та/або лише якісно. Глибше уявлення про точну молекулярну динаміку реплікації HCV покращить, наприклад, розуміння взаємодії вірусних білків (протеїнів), реплікація вірусної РНК (рибонуклеїнова кислота -- одна із трьох основних макромолекул, дві інші -- ДНК та білки) (vRNA) та клітинний метаболізм. Ці знання, швидше за все, знадобляться щоб покращити наше розуміння впливу прямих противірусних засобів (DAA) на взаємодію vRNA та вірусні білки. Крім того, таке розуміння може бути використане для виявлення глобальних механізмів, що беруть участь у реплікації споріднених вірусів.

За останні роки математичні описи динаміки біофізичних процесів стали точнішими. У разі динаміки вірусів компартаментні кінетичні моделі на основі звичайних диференціальних рівнянь (ОДР) тепер представляють сучасний рівень техніки [5–18]. Фактично такі ОДЕ представляють гомогенізовані описи процесів, для яких не враховується просторова роздільна здатність. Для багатьох процесів такі гомогенізації виправдані та корисні. (У цьому дослідженні термін «гомогенізація» стосується усереднення динаміки по всьому простору, а не до інших значень, які з'являються в інших контекстах). Отже, моделі динаміки ВГС на основі ODE дозволили створити нові і важливі ідеї. PDE надають можливість описувати такі ефекти, як просторовий розподіл агентів, що супроводжуються еволюцією vRNA, які неможливо вирішити інакше, і врахувати вплив геометрії на просторово-часову еволюцію невідомого.

7.5.2 Біологічна основа: цикл реплікації вірусного геному

Вірус гепатиту С належить до сімейства Flaviviridae так званої вірусної плюс-ланцюгової РНК [3]. Головна властивість таких вірусів полягає у тому, що вони сильно паразитують на механізмі клітини-хазяїна щоб вплинути на їх власний реплікативний цикл: вРНК, після звільнення всередині клітини від її захисної білкової оболонки (так званий капсид), прилягає безпосередньо до рибосом, які синтезують вірус поліпротеїни. Поліпротеніаза згодом розщеплюється протеазами хазяїна та вірусу. Отримані неструктурні білки (NSP) функціонують для реплікації вірусного геному, тоді як структурні білки (SPs) слугують для інкапсуляції щойно реплікованої вРНК з утворенням віріонів нащадків, які вивільняються з клітини з метою інфікування інших клітин.

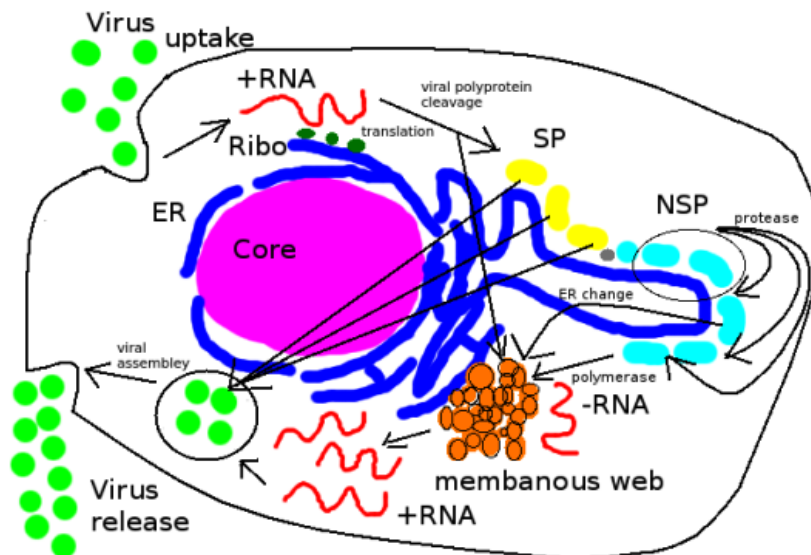


Рис. 7.4. Цикл реплікації вірусу HCV: ендоцитоз вірусу, розкриття вРНК, трансляція вРНК поліпротеїну, розщеплення поліпротеїну на *структурні та неструктурні* білки (SP/NSP), NSP створюють мембранна мережа в ER, реплікація вРНК всередині мережі, нова вРНК і SP, зібрані з новим вірусом частинок, екзоцитоз повністю нових вірусів, які вражають інші клітини

Endoplasmatic Reticulum (ER). Ендоплазматичний ретикулум (EP) — це система мембранних цистерн (сплюснених мішків), що простягаються по всій цитоплазмі. Ендоплазматичний ретикулум можна класифікувати за двома функціонально відмінними формами: гладкий ендоплазматичний ретикулум (SER) і шорсткий ендоплазматичний ретикулум (RER). Морфологічна відмінність між ними полягає у наявності білок-синтезуючих частинок, званих *рибосомами*, прикріплених до зовнішньої поверхні RER.

Моделі, які ми тут представляємо, відтворюють центральні елементи циклу відтворення vRNA HCV: vRNA дифундує від свого початкового місця, де вона була звільнена від вірусної частинки, і прикріплюється до рибосом. На рибосомах виробляються NSP, які дифундують, утворюючи мембрану перетинки на поверхнях ER. Потім вРНК дифундує до мембранних тканин і реплікується. Нова vRNA дифундує до рибосом, де знову починається той самий цикл. Оскільки ми використовуємо реконструйовану геометрію ER і реконструйовані ділянки мембранозної тканини на основі експериментальних даних імунофарбованих клітин тканинної культури, отриманих з гепатоцитів, поверхня ER та області, де розташовані мембранні тканини росту, відповідають їх розташуванням у використаних експериментальних даних.

7.5.3 Динаміка гепатиту С з оптимальним лікуванням

Модель для монотерапії IFN

Розглянемо модель, яка описує динаміку неінфікованих та інфікованих гепатоцитів, а також ВГС і виражається як система трьох звичайних диференціальних рівнянь:

$$\begin{aligned}
\frac{dT}{dt} &= s + rT \left(1 - \frac{T+I}{T_{\max}} \right) - dT - \beta TV, \\
\frac{dI}{dt} &= \beta TV + rI \left(1 - \frac{T+I}{T_{\max}} \right) - \delta I, \\
\frac{dV}{dt} &= (1 - \varepsilon_p) pI - cV.
\end{aligned} \tag{122}$$

Неінфіковані гепатоцити (T) виробляються зі швидкістю s і логістично розмножуються зі швидкістю r , що супроводжується природною смертністю d . Гомеостатичний механізм печінки включається через $T_{\max}$, що є максимальною кількістю гепатоцитів у печінці. HCV (V) інфікує гепатоцити зі швидкістю β , таким чином продукуючи інфіковані гепатоцити (I). Також передбачається, що інфіковані гепатоцити розмножуються логістично зі швидкістю r і мають природний рівень смертності d . За відсутності будь-якого лікування інфіковані гепатоцити продукують HCV зі швидкістю p , яка має швидкість очищення c . Продукція HCV знижується на коефіцієнт $(1 - \varepsilon_p)$ при введенні лікування IFN, де ε_p – ефективність IFN. Модель виключає роль IFN у блокуванні інфекції, оскільки було виявлено, що у цьому випадку він має мінімальний ефект [1,8]. Ця модель допускає два стійких стани [9]:

незаражений стан,

$$T^{(u)} = \frac{T_{\max}}{2r} \left[r - d + \sqrt{(r-d)^2 + \frac{4sr}{T_{\max}}} \right], \quad I^{(u)} = 0, \quad V^{(u)} = 0$$

та заражений стан,

$$T^{(i)} = \frac{1}{2} \left[-D + \sqrt{D^2 + \frac{4sT_{\max}}{rA^2}} \right], \quad I^{(i)} = T^{(i)}(A-1) + T_{\max} - B, \quad V^{(i)} = \frac{1}{c}(1 - \varepsilon_p)pI^{(i)},$$

де

$$A = \frac{1 - \varepsilon_p}{cr} p\beta T_{\max}, \quad B = \frac{\delta T_{\max}}{r}, \quad D = \frac{1}{A} \left[T_{\max} + \frac{dB}{\delta A} - B(1 + 1/A) \right].$$

За фізіологічних умов $r > d$ і $s \leq dT_{\max}$ існує транскритична біфуркація при

$$(1 - \varepsilon_p) = \frac{c(\delta T_{\max} + rT^{(u)} - rT_{\max})}{p\beta T^{(u)}T_{\max}} = C_1^*.$$

Іншими словами, неінфікований стаціонарний стан є стабільним, якщо $(1 - \varepsilon_p) < C_1^*$, тоді як інфікований стаціонарний стан є стабільним, якщо $(1 - \varepsilon_p) > C_1^*$.

Задача оптимального керування

Мета – сформулювати функціонал, який знижує рівні HCV (V) та інфікованих гепатоцитів (I) під час і в кінці терапії. Проблема також вимагає мінімізації терапевтичних побічних ефектів шляхом покарання за надмірне використання IFN.

Цільовий функціонал, який потрібно мінімізувати, визначається як

$$J = \frac{1}{2} [S_{22}I^2(t_f) + S_{33}V^2(t_f)] + \frac{1}{2} \int_{t_i}^{t_f} [Q_{22}I^2(t) + Q_{33}V^2(t) + R\varepsilon_p^2(t)] dt,$$

де S_{ii}, Q_{ii}, R -- коефіцієнти витрат для відповідних змінних.

Цей функціонал вартості обмежує зміни в інтегральних та термінальних концентраціях інфікованих гепатоцитів і віріонів, а також ефективність від нуля. Включення коефіцієнтів мотивується фундаментальною терапевтичною метою зниження рівня концентрації як інфікованих гепатоцитів, так і ВГС. Коефіцієнти S_{ii} мінімізують термінальні рівні I і V , тоді як Q_{ii} сприяють прискореній вірусологічній відповіді – обидва вони досягаються завдяки вищій ефективності. R відображає фінансові та фізіологічні витрати, пов'язані з лікуванням, і покарання за високі рівні ефективності. Точніше, R враховує фінансові обмеження (такі як вартість лікування) і фізіологічні наслідки (такі як побічні ефекти, токсичність і дискомфорт). Функціонал вартості фіксує баланс між швидкістю одужання та терапевтичними побічними ефектами. Вища ефективність, викликаючи швидшу відповідь на терапію, обмежена фізіологічними та фінансовими витратами, спричиненими такою терапією, тому конкуруючий характер умов у функціоналі вартості становить допустимий компроміс між швидкістю терапевтичної відповіді та пов'язані з цим витрати.

Для зручності позначень ми визначаємо змінні стану як $x_1 = T$, $x_2 = I$, $x_3 = V$ і керуючу змінну як $u = \varepsilon_p$. Рівняння стану:

$$\begin{aligned}\frac{dx_1}{dt} &= s + rx_1 \left(1 - \frac{x_1 + x_2}{T_{\max}}\right) - dx_1 - \beta x_1 x_3, \\ \frac{dx_2}{dt} &= \beta x_1 x_3 + rx_2 \left(1 - \frac{x_1 + x_2}{T_{\max}}\right) - \delta x_2, \\ \frac{dx_3}{dt} &= (12 - u)px_2 - cx_3.\end{aligned}\tag{123}$$

Визначивши вектор стану $x = (x_1, x_2, x_3)^T$, наведені вище рівняння можна записати як $\dot{x} = f(x(t), u(t), t)$ із початковою умовою $x(t_i) = x_0$. Вищевказану проблему можна переформулювати як частину загальнішого налаштування, як описано нижче:

Знайти допустиме оптимальне управління $u^*: [t_i, t_f] \rightarrow [\underline{u}, \bar{u}]$, яке для системи $\frac{dx}{dt} = f(x(t), u(t), t)$, $x(t_i) = x_0$ мінімізує функціонал вартості

$$\begin{aligned}\mathcal{J} &= \frac{1}{2} x^T(t_f) S x(t_f) + \int_{t_i}^{t_f} \frac{1}{2} [x^T(t) Q x(t) + u^T(t) R u(t)] dt \\ &= \mathcal{K}(x(t_f), t_f) + \int_{t_i}^{t_f} \mathcal{L}(x(t), u(t), t) dt.\end{aligned}$$

Гамільтоніан, визначений за допомогою динамічного обмеження $f(x(t), u(t), t)$ і лагранжіан $\mathcal{L}(x(t), u(t), t)$ через приєднаний вектор λ^T (з тією ж розмірністю, що й вектор стану), виглядає як

$$\mathcal{H}(x(t), u(t), t) = \mathcal{L}(x(t), u(t), t) + \lambda^T(t) f(x(t), u(t), t).$$

Використовуючи принцип мінімуму Понтрягіна, необхідні умови (в термінах гамільтоніана) для того, щоб u^* було оптимальним керуванням:

$$\frac{dx^*(t)}{dt} = \nabla_{\lambda} \mathcal{H}(x^*(t), u^*(t), \lambda^*(t), t) = f(x^*(t), u^*(t), t);$$

$$x^*(t) = x_0.$$

$$\begin{aligned} \frac{d\lambda^*(t)}{dt} &= -[\nabla_x \mathcal{H}(x^*(t), u^*(t), \lambda^*(t), t)]^T \\ &= -[\nabla_x \mathcal{L}(x^*(t), u^*(t), t)]^T - [\nabla_x (f(x^*(t), u^*(t), t))]^T \lambda^*(t); \\ \lambda^*(t_f) &= \nabla_x \mathcal{K}(x^*(t_f), t_f) = Sx^*(t_f); \end{aligned}$$

$$\nabla_u \mathcal{H}(x^*(t), u^*(t), \lambda^*(t), t) = \nabla_u \mathcal{L}(x^*(t), u^*(t), t) + \lambda(t) \nabla_u f(x^*(t), u^*(t), t) = 0.$$

Оптимальна система керування, таким чином, є прямим рівнянням стану та зворотним зв'язаним рівнянням разом із регулярним управлінням. Ця проблема, будучи нелінійною та пов'язаною за своєю природою, потребує вирішення за допомогою паралельної та ітераційної чисельної процедури.

1. Апроксимація починається з початкового припущення для елемента керування $u(t) = u^{(0)}(t)$.

2. За допомогою цього керування $u^{(0)}(t)$, $x^{(0)}(t)$ знаходять чисельне розв'язання $\frac{dx(t)}{dt} = f(x(t), u^{(0)}(t), t)$, $x(t_i) = x_0$, яке потім використовують для вирішення

$$\frac{d\lambda(t)}{dt} = -[\nabla_x \mathcal{H}(x^{(0)}(t), u^{(0)}(t), \lambda(t), t)], \quad \lambda(t_f) = Sx^{(0)}(t_f)$$

для визначення $\lambda^{(0)}(t)$.

3. Для отримання оновленого елемента керування оцінюється градієнт

$$\nabla_u \mathcal{H}(x^{(0)}(t), u^{(0)}(t), \lambda^{(0)}(t), t):$$

$$u^{(1)}(t) = u^{(0)}(t) - \tau(0) \nabla_u \mathcal{H}(x^{(0)}(t), u^{(0)}(t), \lambda^{(0)}(t), t).$$

Далі маємо ітераційну процедуру

$$u^{(k+1)}(t) = u^{(k)}(t) - \tau(0) \nabla_u \mathcal{H}(x^{(k)}(t), u^{(k)}(t), \lambda^{(k)}(t), t).$$

Ця процедура повторюється доти, доки не буде задоволений певний критерій збіжності, наприклад

$$\|u^{(k+1)}(t) - u^{(k)}(t)\| < \delta \quad \Delta \mathcal{J}^{(k)} / \mathcal{J}^{(k)} < \delta.$$

Оптимальне управління моделювалося на періоді 50 діб. Коефіцієнти $S_{22} = S_{33} = 10^{-5}$, $Q_{22} = Q_{33} = 10^{-5}$, $R = 5000$.

Модель комбінованого лікування IFN і рибавірином

Розглядається розширена модель, яка включає терапевтичні ефекти рибавіріну до моделі. Ця модель для комбінованого лікування IFN і рибавірином складається з чотирьох пов'язаних ODE:

$$\begin{aligned}
\frac{dT}{dt} &= s + rT \left(1 - \frac{T+I}{T_{\max}} \right) - dT - \beta TV_I, \\
\frac{dI}{dt} &= \beta TV_I + rI \left(1 - \frac{T+I}{T_{\max}} \right) - \delta I, \\
\frac{dV_I}{dt} &= (1-\rho)(1-\varepsilon_p) pI - cV_I, \\
\frac{dV_{NI}}{dt} &= \rho(1-\varepsilon_p) pI - cV_{NI}.
\end{aligned} \tag{124}$$

Рибавірін діє, перетворюючи частину ρ новоутворених віріонів на неінфекційні, що ділить популяцію віріонів на дві різні популяції віріонів, а саме інфекційну (V_I) і неінфекційну (V_{NI}). Ця модель також допускає два стійкі стани [9]:

неінфікований стан

$$T^{(u)} = \frac{T_{\max}}{2r} \left[r - d + \sqrt{(r-d)^2 + \frac{4rs}{T_{\max}}} \right], I^{(u)} = 0, V_I^{(u)} = 0, V_{NI}^{(u)} = 0.$$

інфікований стан

$$\begin{aligned}
T^{(i)} &= \frac{1}{2} \left[-D + \sqrt{D^2 + \frac{4sT_{\max}}{rA^2}} \right], I^{(i)} = T^{(i)}(A-1) + T_{\max} - R, \\
V_I^{(i)} &= \frac{1}{c} [(1-\varepsilon_p)(1-\rho)pI^{(i)}], V_{NI}^{(i)} = \frac{\rho V_I^{(i)}}{1-\rho},
\end{aligned}$$

де

$$A = \frac{1}{cr} [(1-\varepsilon_p)(1-\rho)p\beta T_{\max}], B = \frac{\delta T_{\max}}{r}, D = \frac{1}{A} \left[T_{\max} + \frac{dB}{\delta A} - B \left(\frac{1}{A} + 1 \right) \right].$$

У фізіологічних умовах $r > d$ і $s \leq dT_{\max}$ існує транскритична біфуркація на

$$(1-\varepsilon_p)(1-\rho) = \frac{c(\delta T_{\max} + rT^{(u)} - rT_{\max})}{p\beta T^{(u)}T_{\max}} = C_2^*.$$

Іншими словами, неінфікований стабільний стан є стабільним, якщо $(1-\varepsilon_p)(1-\rho) < C_2^*$, тоді як інфікований стабільний стан є стабільним, якщо $(1-\varepsilon_p)(1-\rho) > C_2^*$.

Задача оптимального керування комбінованим лікуванням

Ми визначаємо модифікований цільовий функціонал для моделі з чотирьох рівнянь, описаної у попередньому розділі. У визначенні функціоналу мети розглядаємо інтегральні значення концентрації інфікованих гепатоцитів (I) та інфекційного вірусного навантаження (V_I), а також кінцеві значення інфікованих гепатоцитів і обох віріонних популяцій (V_I , V_{NI}). Інтегральна величина неінфекційного вірусного навантаження не враховується (тобто має нульовий коефіцієнт витрат), оскільки він не є шкідливий під час курсу лікування. Функціонал також містить обмеження побічних ефектів, що є результатом використання обох IFN і рибавіріну, і забезпечує конкуруючий механізм. Крім того, функціонал інкапсулює потенційний компроміс між швидким терапевтичним

ефектом і низькими побічними ефектами. Більш висока ефективність ліків полегшує захворювання швидше, однак більш висока ефективність є демотивованою через більш високі фізіологічні та фінансові витрати, описані раніше. З цих біомедичних міркувань функціонал мети для комбінованої терапії визначається як

$$\mathcal{J} = \frac{1}{2} \left[S_{22} I^2(t_f) + S_{33} V_I^2(t_f) + S_{44} V_{NI}^2(t_f) \right] + \frac{1}{2} \int_{t_i}^{t_f} \left[Q_{22} I^2(t) + Q_{33} V_I^2(t) + R_{11} \varepsilon_p^2(t) + R_{22} \rho^2(t) \right] dt,$$

Вектор стану для цієї моделі визначається як $x = [x_1, x_2, x_3, x_4]^T$, де $x_1 = T, x_2 = I, x_3 = V_I, x_4 = V_{NI}$. Аналогічно, вектор керування визначається як $u = [u_1, u_2]^T$, $u_1 = \varepsilon_p, u_2 = \rho$. Система, що описана рівнянням ((124)), записується у вигляді

$$\frac{dx}{dt} = f(x(t), u(t), t), \quad x(t_f) = x_0.$$

Подальша математична та чисельна обробка задачі аналогічна підходу, прийнятому для моделі монообробки.

Початкові значення, як й у випадку першої моделі, обираються як $(4, 4 \cdot 10^6; 1, 4 \cdot 10^4; 1, 3 \cdot 10^6; 0)$. $S_{22} = S_{33} = S_{44} = 10^{-5}$; $Q_{22} = Q_{33} = 10^{-5}$; $R_{11} = R_{22} = 5000$.

7.5.4 Моделі PDE

Математичний опис руху vRNA, трансляції та розщеплення поліпротеїнів, рух NSP та накопичення/ зв'язування в ділянках білка, копіювання vRNA та рух нової vРНК до рибосомальних областей для вироблення нових NSP формулюються шляхом введення концентрацій:

- вРНК: $R(x, t)$;
- вірусного поліпротеїну: $P(x, t)$;
- Web-білка (тобто NSP, який виникає з розщепленого поліпротеїну та який накопичується для формування веб-областей) на попередньо визначених мережах: $Q(x, t)$;
- фактора господаря: $H(x, t)$.

Отже, просторово-часова оцінка R, P, W і H моделюється за допомогою нелінійно зв'язаних поверхневих PDE (sPDE) ($x \in D \subset R^3$):

$$\begin{aligned} \frac{\partial R}{\partial t} &= \operatorname{div}_T(D_R \nabla_T R) + r_1 W H R, \\ \frac{\partial P}{\partial t} &= \operatorname{div}_T(D_P \nabla_T P) + r_2 R - r_3 P, \\ \frac{\partial W}{\partial t} &= \operatorname{div}_T(D_W \nabla_T W) + r_3 P, \\ \frac{\partial H}{\partial t} &= \operatorname{div}_T(D_H \nabla_T H) - r_4 W H R, \end{aligned} \tag{125}$$

де D_R, D_P, D_W, D_H -- кусково постійні коефіцієнти дифузії. Ми вважаємо, що коефіцієнти дифузії є кусково постійними, щоб забезпечити дифузію у тих областях, в яких біологічно підтримується, що речовина рухається і для того, щоб заборонити дифузію у тих областях,

в яких не допускається рух речовини. Наприклад, поліпротеїнам не дозволяється дифундувати на поверхні ER подалі від рибосом, веб-білкам не дозволяється дифундувати подалі від ділянок павутини, ні на поверхні ER, ні в цитозол.

Ми використовуємо такі позначення: тангенціальну дивергенцію позначаємо як div_T , а ∇_T — це оператор тангенціального градієнта. Закони дифузійної реакції (106) є поверхневими диференціальними рівняннями із частинними похідними (sPDE). Вони описують процеси дифузії та реакції, що відбуваються на мембрані ER і поверхонь полотна. Швидкості реакції $r_1 \div r_4$ є кусково постійні функції.

$$r_i(x) > 0, x \in W(R). \quad (126)$$

Біологічний зміст реакційних констант полягає в такому:

- r_1 описує швидкість полімеризації нової vRNA в термінах багатолінійних концентрацій vPHK, веб-білка та фактора господаря;
- r_2 описує швидкість трансляції поліпротеїну в термінах лінійної концентрації vRNA;
- r_3 описує швидкість розщеплення поліпротеїну на павутинний (акумуючий) білок;
- r_4 описує швидкість виснаження фактора господаря під час полімеризації vPHK у термінах багатолінійної концентрації vRNA, веб-білка та фактора господаря.

Зверніть увагу, що vRNA полімеризація і виснаження фактора господаря пропорційні з єдиною різницею у відношенні від r_1 до r_4 , щоб забезпечити різну кількість виснаження фактора господаря порівняно з полімеризацією vRNA, тоді як vRNA копіюється.

Фактор хазяїна Н вводився, щоб уникнути нереального виробництва vRNA, яке б інакше розходилося. Фактор господаря споживається паралельно, коли vPHK реплікується в мережі. Більше того, оскільки реплікація вірусу залежить від можливості порушити метаболізм хазяїна (фактично, реплікація не може відбутися, якщо відсутній фактор господаря), введення фактора господаря необхідно для того, щоб модель була реалістичною.

Щоб змодельовати той факт, що поліпротеїн не дифундує за межі R, тоді як веб-білок не дифундує за межі $W \cup R$, вибираємо D_p і D_w як

$$D_p(x) > 0, \forall x \in R. \quad (127)$$

$$D_w(x) > 0, \forall x \in W \cup R. \quad (128)$$

Крім того, прописуємо початкові умови

$$R(x, 0) = \begin{cases} R_0 & \in R_{i_*} \ i_* = \overline{1, 7}, \\ 0 & \in D \setminus R_{i_*}, \end{cases} \quad (129)$$

$$W(x, 0) = 0 \in D, \quad (130)$$

$$H(x, 0) = 1 \in D, \quad (131)$$

$$P(x, 0) = 0 \in D. \quad (132)$$

Усі поверхневі PDE у цій статті розв'язуються за нульових межових умов Неймана на ∂D для кожної з концентрацій.

Оскільки параметри D_p , r_2 і r_3 такі, що поліпротеїн дуже швидко розпадається (тобто розщеплюється) порівняно з іншими видами, асимптотична поведінка системи

визначається як

$$\begin{aligned}\frac{\partial R}{\partial t} &= \operatorname{div}_T(D_R \nabla_T R) + r_1 RWH, \\ \frac{\partial W}{\partial t} &= \operatorname{div}_T(D_W \nabla_T W) + r_2 R, \\ \frac{\partial H}{\partial t} &= \operatorname{div}_T(D_H \nabla_T H) - r_4 RWH,\end{aligned}\tag{133}$$

Зазначимо, що реакція $r_1 RWH$ є багатолінійна функція R, W і H . Проте це наближення можливе доти, доки концентрації R, W і H досить малі.

Запишемо (скалярну) фізичну величину $f: \mathcal{D} \rightarrow \mathbf{R}$ і непорожню множину $Y \rightarrow \mathcal{D}$ як $[f]_Y$, щоб вказати обмеження f на множину, в якій вона відмінна від нуля. Це зроблено для чіткішої візуалізації, коли потрібно оцінити f . Отже, рівняння (114) стають

$$\begin{aligned}\frac{\partial R}{\partial t} &= \operatorname{div}_T(D_R \nabla_T R) + [r_1 RWH]_W, \\ \frac{\partial W}{\partial t} &= \operatorname{div}_T(D_W \nabla_T W)_{W \cup R} + [r_2 R]_R, \\ \frac{\partial H}{\partial t} &= \operatorname{div}_T(D_H \nabla_T H) - [r_4 RWH]_W,\end{aligned}\tag{134}$$

Крім того, оскільки ми вибираємо кусково сталі коефіцієнти дифузії, поверхнева дифузія керується оператором Лапласа-Бельтрамі Δ_T , тобто проекцією оператора Лапласа на тангенціальний простір двовимірної ER-гіперповерхні $\mathcal{E}$, яка вбудована в повний 3D простір [49]. Отже, рівняння можна переписати у простому вигляді

$$\begin{aligned}\frac{\partial R}{\partial t} &= D_R \Delta_T R + [r_1 RWH]_W, \\ \frac{\partial W}{\partial t} &= [D_W \Delta_T W]_{W \cup R} + [r_2 R]_R, \\ \frac{\partial H}{\partial t} &= D_H \nabla_T H - [r_4 RWH]_W,\end{aligned}\tag{135}$$

Вихідні рівняння (106), що включають проміжний поліпротеїновий стан, можна переписати як

$$\begin{aligned}\frac{\partial R}{\partial t} &= D_R \Delta_T R + [r_1 RWH]_W, \\ \frac{\partial P}{\partial t} &= [D_P \Delta_T P + r_2 R - r_3 P]_R, \\ \frac{\partial W}{\partial t} &= [D_W \Delta_T W]_{W \cup R} + [r_3 R]_R, \\ \frac{\partial H}{\partial t} &= D_H \nabla_T H - [r_4 RWH]_W,\end{aligned}\tag{136}$$

Значення параметрів: $D_R = 0,3 (\mu m)^2 / s$, $D_P = 0,3 (\mu m)^2 / s$, $D_W = 0,5 (\mu m)^2 / s$,

$D_H = 0,05 (\mu m)^2/s$; $r_1 = 15 1/s$, $r_2 = 100 1/s$, $r_3 = 90 1/s$, $r_4 = 1,5 1/s$; $R_0 = 10 1/(\mu m)^2$, $webs = 7$.

Початкові та межові умови

$$R(x,0) = \begin{cases} R_0, & \forall x \in S_V, \\ 0, & \text{else} \in \Omega \end{cases} \quad (137)$$

$$N(x,0) = W(x,0) = 0 \forall x \in \Omega, \quad (138)$$

$$H(x,0) = H_0 \forall x \in \Omega. \quad (139)$$

$$H(x,t) = H_0 \forall x \in \partial\Omega \wedge \forall t. \quad (140)$$

де S_V -- маленька сфера в цитозолі.

Решта межових умов є однорідні умови Неймана і усі концентрації (крім vRNA) є нульові на початку.

7.5.5 Розширення моделі sPDE

Модель sPDE у рівняннях має згадані раніше обмеження. Як і в існуючій моделі на основі ОДУ, член реакції $r_1 WHR$ є багатолінійна функція R , W і H . Це може бути правдоподібним для досить малих концентрацій R , W і H , але тепер замінено на більш реалістичні умови.

Ми використовуємо умови реакції, залежні від концентрації, у рівняннях щоб забезпечити більшу точність цих величин. Зокрема, нові члени реакції прагнуть до нуля для обнулення концентрації, і вони наближаються до постійних значень при зростанні концентрацій. Тому члени реакції замінені на

$$\hat{r}_1(W, H)R = \bar{r}_1 \frac{H^n}{H^n + H_0^n} \frac{W^m}{W^m + W_0^m} R, \quad (141)$$

$$\hat{r}_4(W, H)R = \bar{r}_4 \frac{H^n}{H^n + H_0^n} \frac{W^m}{W^m + W_0^m} R, \quad (142)$$

з константами $\bar{r}_1$, $\bar{r}_4$, причому показники n та m є натуральними числами, а константи H_0 , W_0 є контрольними значеннями фактора хазіяїна та концентрації веб-білка відповідно.

Тоді рівняння замінюються на

$$\frac{\partial R}{\partial t} = \text{div}_T(D_R \nabla_T R) + [\hat{r}_1 R]_W, \quad (143)$$

$$\frac{\partial P}{\partial t} = [\text{div}_T(D_P \nabla_T P)]_R + [r_2 R - r_3 P]_R, \quad (144)$$

$$\frac{\partial W}{\partial t} = [\text{div}_T(D_W \nabla_T W)]_{W \cup R} + [r_3 P]_R, \quad (145)$$

$$\frac{\partial H}{\partial t} = \text{div}_T(D_H \nabla_T H) - [r_4 R]_W, \quad (146)$$

З рівнянь (122),(123) випливає, що члени реакції $\hat{r}_1(W, H)$, $\hat{r}_4(W, H)$ дорівнюють нулю коли H або W дорівнює нулю. Такі умови мають місце, наприклад, у випадку виснаження фактору господаря, описаний у рівнянні (127). Далі ці вирази прагнуть до сталих значень $\bar{r}_1$, $\bar{r}_4$ відповідно за $H_0/H \rightarrow 0$, $W_0/W \rightarrow 0$. Цей процес не залежить від H ,

коли відношення $H_0^n/H^n \rightarrow 0$, що означає, що швидкість виснаження не залежить від H , коли $H \gg H_0$.

Крім того, нова модель розрізняє різні стани компонентів, наприклад, розрізняє РНК ВГС, зв'язану з рибосомами, і РНК ВГС, яка реплікується всередині павутин. На початку процесу реплікації РНК ВГС передбачається, що РНК ВГС зв'язується в одній рибосомальній області R_i і не може дифундувати за її межі. Більше того, це РНК HCV змушує рибосоми транслювати поліпротеїн, якому також дозволяється дифундувати лише всередині рибосомальної області. Концентрація РНК HCV, зв'язаної з рибосомами, позначається R_r і концентрація поліпротеїну описується P . Поліпротеїн розщеплюється на два вірусних білка, а саме веб-білок із концентрацією W та інший білок (наприклад, неструктурний білок 5a, NS5a) з концентрацією N . Веб-білку дозволяється дифундувати в рибосомній області та геометрично визначеній підобласті W_i MW і формує MW в межах цієї експериментально визначеної області. Натомість неструктурний білок NS5a може вільно дифундувати поверхнею ER поверхні $\mathcal{E}$ та рибосомні ділянки R .

Коефіцієнти дифузії моделюються таким чином, що вони залежать не тільки від підобласті, але також від концентрації вірусних компонентів і фактора господаря. Отже, рівняння дифузії–реакції (рівняння (124)–(127) переписуються як

$$\frac{\partial R_r}{\partial t} = [\operatorname{div}_T(D_{R_r} \nabla_T R_r)] R + \left[\tilde{r}_1 R_f \frac{H^i}{H^i + F_1} - \tilde{r}_2 R_r \frac{R_r^p}{R_r^p + F_0} \frac{W^j}{W^j + F_2} \frac{H^k}{H^k + F_3} \right] R, \quad (147)$$

$$\frac{\partial P}{\partial t} = [\operatorname{div}_T(D_P \nabla_T P)] R + \left[\tilde{r}_4 R_r \frac{R_r^p}{R_r^p + F_0} \frac{H^q}{H^q + F_6} \right] R - [\tilde{r}_5 P] R, \quad (148)$$

$$\frac{\partial N}{\partial t} = [\operatorname{div}_T(D_N \nabla_T N)] R \cup \mathcal{E} + [\tilde{r}_5 P] R, \quad (149)$$

$$\frac{\partial W}{\partial t} = [\operatorname{div}_T(D_W \nabla_T W)] R \cup R + \left[\tilde{r}_5 P - v \tilde{r}_2 R_r \frac{R_r^p}{R_r^p + F_0} \frac{W^j}{W^j + F_2} \frac{H^k}{H^k + F_3} \right] R, \quad (150)$$

$$\frac{\partial C}{\partial t} = [\operatorname{div}_T(D_C \nabla_T C)] R \cup R + \left[\tilde{r}_2 R_r \frac{R_r^p}{R_r^p + F_0} \frac{W^j}{W^j + F_2} \frac{H^k}{H^k + F_3} \right] R, \quad (151)$$

$$\frac{\partial R_f}{\partial t} = \operatorname{div}_T(D_{R_f} \nabla_T R_f) + \left[\tilde{r}_6 C \frac{W^n}{W^n + F_5} \frac{H^m}{H^m + F_4} \right] W - \left[\tilde{r}_1 R_f \frac{H^i}{H^i + F_1} \right] R, \quad (152)$$

$$\frac{\partial H}{\partial t} = \operatorname{div}_T(D_H \nabla_T H) - \left[\tilde{r}_3 C \frac{W^n}{W^n + F_5} \frac{H^m}{H^m + F_4} \right] W. \quad (153)$$

Параметри та коефіцієнти цієї системи мають такі характеристики: Показники степеню i, j, k, m, n, p, q — натуральні числа, $\tilde{r}_1, \dots, \tilde{r}_6$ — дійсні константи $[1/c]$ і являють собою характерні значення членів реакції у рівняннях (128)–(134). Параметр $v > 1$ є константа, яку запроваджено, щоб підкреслити той факт, що швидкість адсорбції веб-білків має бути більша за ту, яка стосується рибосомальної РНК ВГС (фактично, $v \gg 1$). Параметри $F_0, \dots, F_6$ всі додатні і, отже, призводять до того, що швидкості реакції мають постійні асимптотичні значення для зростання значення концентрації видів. За

коефіцієнти дифузії D_{R_f} , D_P , D_N , D_W прийнято константи, тоді як D_C , D_{R_f} і D_H залежать від концентрації і визначаються як

$$D_C := \hat{D}_C(W) = D_C^{(0)} \frac{W}{W + F_7}, \quad (154)$$

$$D_{R_f} := \hat{D}_{R_f}(W, N) = D_{R_f}^{(0)} \left[\mathcal{F} \frac{N}{N + F_8} + \frac{W}{W + F_9} \right], \quad (155)$$

$$D_H := \hat{D}_H(W, N) = D_H^{(0)} \left[\mathcal{G} \frac{N}{N + F_{10}} + \frac{W}{W + F_{11}} \right]. \quad (156)$$

$D(0)_C$, $D(0)_{R_f}$, $D(0)_H$ – постійні коефіцієнти дифузії, а $\mathcal{F}$ і $\mathcal{G}$ – постійні коефіцієнти дифузії налаштування параметрів моделі впливів доданків $N/(N + F_8)$ і $N/(N + F_{10})$ на D_{R_f} і D_H відповідно, а $F_7, \dots, F_{11}$ мають еквівалентні значення, аналогічні $F_1, \dots, F_6$. Рівняння (128)–(134) розв’язуються із урахуванням нульової межевої умови Неймана на ∂D та початкової умови

$$R_r(x, t_0) = \begin{cases} R_{r0}, & \forall x \in \mathbf{R}, i_* = \overline{1, 7}, \\ 0, & \forall x \in D \setminus \mathbf{R}, \end{cases} \quad (157)$$

$$H(x, t_0) = H_0, \forall x \in D, \quad (158)$$

Решта початкових умов нульові.

7.5.6 Додаток

Нехай $\omega \subset \mathcal{S}$ — відкрита множина тривимірного евклідова простору $\mathcal{S}$ і нехай $f : \omega \rightarrow \mathbf{R}$ - скалярне поле, визначене над ω . Якщо всі частинні похідні від f існують всюди в ω , градієнт (“grad” або ∇) від f у заданій точці $x \in \Omega$ визначається як

$$\text{grad} f(x) \equiv \nabla f(x) = \sum_{k=1}^3 \frac{\partial f(x)}{\partial x_k} e_k, \quad (159)$$

де $\{e_k\}_{k=1}^3$ — базисні вектори, а трійка дійсних чисел (x_1, x_2, x_3) позначає координати $x \in \Omega$.

Крім того, векторне поле u над $\Omega \subset \mathbf{R}^3$ розкладається як

$$u = u_1 e_1 + u_2 e_2 + u_3 e_3, \quad (160)$$

трійка дійсних функцій (u_1, u_2, u_3) представляє компоненти u в декартовому векторному базисі $\{e_k\}_{k=1}^3$. Якщо кожна функція u_k диференційовна,

$$\text{grad} u(x) \equiv \nabla u(x) = \sum_{k,j=1}^3 \frac{\partial u_k}{\partial x_j} e_k \otimes e_j, \quad (161)$$

де символ “ $\otimes$ ” позначає діадичний добуток між векторами. Нагадуємо, що для будь-якої пари векторів v, w діадичний добуток $v \otimes w$ є тензором другого порядку з компонентами $[v \otimes w]_{k,j} = v_k w_j$. Нарешті, розбіжність u отримується шляхом зведення двох індексів $\text{grad} u$ [65], тобто

$$\operatorname{div} u(x) = \frac{\partial u_1(x)}{\partial x_1} + \frac{\partial u_2(x)}{\partial x_2} + \frac{\partial u_3(x)}{\partial x_3} = \sum_{k=1}^3 \frac{\partial u_k(x)}{\partial x_k}. \quad (162)$$

Якщо $\operatorname{grad} u(x)$ існує, $\operatorname{div} u(x)$ можна обчислити як слід матричного представлення $\operatorname{grad} u(x)$.

У своїй роботі ми часто використовуємо оператори градієнта та дивергенції, оскільки обидва оператори у рівняннях дифузії–реакції, на яких базується наша модель. Фактично кожне з таких рівнянь представляє закон балансу маси для даного виду моделі і може бути записаний як

$$\frac{\partial c}{\partial t} = -\operatorname{div} q + r, \quad (163)$$

де c – концентрація речовини, q – вектор потоку маси, що описує рух виду, і r є відповідне джерело або поглинач.

Рівняння (144) є локальна форма закону балансу маси і походить від інтегральної форми

$$\frac{d}{dt} \int_P c dv = - \int_{\partial P} q \cdot n da + \int_P r dv, \quad (164)$$

У рівнянні (145) $P \subseteq \Omega$ є загальна частина Ω , ∂P є його межа, n – вектор нормалі до ∂P , dv та da позначають міри об'єму та поверхні відповідно. Фізичне (і біологічне) значення рівняння (145) полягає у тому, що маса виду, що міститься в P з концентрацією c змінюється з часом у відповідь на потік виду через ∂P , і на його виробництво або виснаження в P . Потік виду, заданий компонентом q вздовж n , визначається скалярним добутком $q \cdot n$. Продовжуючи q всередину P та припускаючи диференційовність, згідно теореми Гауса рівняння (145) набуває вигляд

$$\int_P \frac{\partial c}{\partial t} dv = \int_P [-\operatorname{div} q + r] dv. \quad (165)$$

Крім того, вимога, щоб рівняння (146) застосовувалося для будь-якого вибору P , дозволяє переписати рівняння (146) згідно з рівнянням (144).

Підсумовуючи, використання розбіжності вектора потоку маси, тобто $\operatorname{div} q$, дозволяє перетворити поверхневий інтеграл від $q \cdot n$ в об'ємний інтеграл від $\operatorname{div} q$, що міститься у рівнянні (146).

Рівняння (144) або інтегральна форма (145) містить два скалярних невідомих, тобто c і r , і векторне невідоме q , яке вводить трійку невідомих скалярних компонент (q_1, q_2, q_3) . Отже, щоб закрити математичну проблему, пов'язану з рівнянням (144), необхідно забезпечити зв'язок між q і r із c , або визначити його із самого початку.

Якщо припустити, що рух розглянутого виду відбувається шляхом дифузії, потрібно описати спонтанний і незворотний процес, за допомогою якого рух видів керується просторовою нерівномірністю його розподілу. Тобто, якщо підготувати експеримент, у якому концентрація виду, c , змінюється у просторі, можна спостерігати рух виду, який прагне відновити незалежну від простору концентрацію. З цієї причини і оскільки дескриптор просторової мінливості c є градієнт концентрації, ∇c , ми визначаємо ∇c як тригер руху виду. Якщо середовище, в якому еволюціонує вид, є ізотропне, найпростіший спосіб зв'язати q із ∇c – це припустити лінійну залежність типу

$q = -D\nabla c$, де D -- коефіцієнт дифузії. Зауважимо, що у фізиці, біології та інженерії можна знайти багато інших законів такого виду: закон молекулярної дифузії, закон теплопровідності Фур'є, закон фільтрації рідини Дарсі через пористі середовища та закон електропровідності Ома.

Вираз q , який використовується у даній роботі, тобто $q = -D\nabla c$, насправді є окремий випадок закону дифузії Фіка. Підставляючи його у рівняння (144), маємо

$$\frac{\partial c}{\partial t} = \operatorname{div}(D\nabla c) + r. \quad (166)$$

У рівнянні (147) член реакції, r , може бути або заданий із самого початку, або пов'язаний із c : $r = \hat{r}(c)$.

Щоб описати дифузію даної речовини на (відкритих) поверхнях, використовується тангенціальний градієнт і оператори тангенціальної дивергенції. Для скалярного поля c тангенціальний градієнт позначається $\nabla_T c$ і визначається як

$$\nabla_T c = [\nabla c]P, \quad (167)$$

де $P = I - n \otimes n$ називають *оператор проекції*, I — тотожний тензор, а n — вектор одиничної нормалі до поверхні S , по якій відбувається дифузія. Щоб рівняння (148) мало сенс, c має бути обмежено S . Отже, c слід розуміти як $c|_S$. Відповідно, тангенціальний вектор потоку маси, пов'язаний із $c|_S$, задається як $q^\sigma = -D\nabla_T c$.

Щоб записати закон балансу маси для S , потрібно взяти тангенціальну розбіжність потоку маси вектору. Загалом, для векторного поля y , визначеного на S , але яке не обов'язково є дотичним до S , тангенціальна розбіжність визначається як [67]

$$\operatorname{div}_T y := \operatorname{tr}(P\nabla_T y) = \operatorname{div}_T y^\sigma - 2\kappa\psi_n, \quad (168)$$

де tr -- слід, $y^\sigma = Py$ — проекція y на S , κ — середня кривизна S , і ψ_n -- компонента y , нормальна до S . Ми підкреслюємо, що для тангенціального векторного поля, як і у випадку q^σ вважається, що $q^\sigma = Pq^\sigma$, і рівняння (149) стає незалежним від середньої кривизни, тим самим зменшуючи до

$$\operatorname{div}_T q^\sigma = -\operatorname{div}_T (D\nabla_T c), \quad (169)$$

Відповідно, рівняння дифузії-реакції (рівняння (147)) стає

$$\frac{\partial c}{\partial t} = \operatorname{div}_T (D\nabla_T c) + r, \quad (170)$$

і є локальна форма інтегрального рівняння

$$\frac{d}{dt} \int_{\mathcal{A}} c da = \int_{\partial \mathcal{A}} (-D\nabla_T c) \nu ds + \int_{\mathcal{A}} r da, \quad (171)$$

де $\mathcal{A} \subseteq S$ — підповерхня S — замкнута регулярна лінія, що міститься в S і визначає межу $\mathcal{A}$, ν — поле одиничних векторів, нормальних до $\partial \mathcal{A}$ та дотичних до $\mathcal{A}$, а ds — міра лінії.

Біологічний сенс використання операторів тангенціального простору полягає у тому, що вони дозволяють обмежувати переміщення компонентів на двовимірну поверхню, вбудовану в 3D, наприклад, для обмеження руху біологічних агентів, таких як білки, на поверхню ER.

7.6 Математичні моделі ВІЛ-інфекції

7.6.1 Вступ

Зараження вірусом імунодефіциту людини 1 типу (ВІЛ-1) має багато кількісних ознак [26]. Математичні моделі ВІЛ-інфекції можуть надати уявлення про динаміку вірусного навантаження *in vivo* і може відігравати значну роль у розвитку кращого розуміння ВІЛ/СНІДу та медикаментозної терапії [23, 37].

Вважається, що ЦТЛ є основним імунним фактором господаря для визначення вірусного навантаження [22]. Коли ВІЛ проникає в організм, він націлений на $CD4^+$ Т-клітини. Ці клітини можна вважати командними центрами імунної системи.

CTL – це клітини, які мають на меті ліквідувати інфекцію шляхом знищення інфікованих клітин [6].

Відомо, що вірус імунодефіциту людини (ВІЛ) інфікує $CD4^+$ Т-хелперні клітини, які є важливою частиною імунної системи, оскільки вони полегшують реакцію організму на багато поширених, але потенційно смертельних інфекцій. Через свою летальність і квазі-невиліковність ВІЛ має стати серйозною проблемою для здоров'я людини, що є наразі відповідальним за мільйони інфекцій і смертей.

Без лікування препаратами від ВІЛ ВІЛ-інфекція прогресує поетапно, з часом погіршується. Фаза первинної інфекції характеризується сильним вірусом реплікації, яка супроводжується відповіддю сильного імунітету. У другій фазі ВІЛ-інфекції інфіковані особи не мають симптомів, але мають стійкий вірус реплікації, що в кінцевому підсумку призводить до розвитку СНІДу [8], який є останньою, найважчою стадією ВІЛ інфекції. Особам діагностують СНІД, якщо вони мають кількість Т-клітин менше 200 клітин/мм^3 або якщо вони мають певні опортуністичні інфекції. Без лікування, люди зі СНІДом зазвичай живуть близько 3 років.

Математичні моделі зробили суттєвий внесок у розуміння ВІЛ-інфекції, імунних відповідей та антиретровірусне лікування. Для вивчення динаміки вірусів у математичні моделі також були включені часові затримки. Відомо, що немає ідеального лікування ВІЛ-інфекції, але ліки від ВІЛ можуть запобігти переходу ВІЛ до СНІДу. Ліки від ВІЛ також знижують ризик зараження ВІЛ передачею іншим людям.

7.6.2 Внутрішньоклітинна ВІЛ-модель із затримкою

Модель розглядає чотири змінні стану: $Z(t)$ -- концентрація неінфікованих клітин, $I(t)$ -- концентрація інфікованих клітин, $V(t)$ -- концентрація вільних вірусних частинок, а $T(t)$ -- концентрація CTL в момент часу t .

Для опису динаміки клітин зроблено такі припущення [22]: неінфіковані клітини виробляються із постійною швидкістю λ і гинуть зі швидкістю mZ . Виробляються інфіковані клітини від неінфікованих клітин і вільних вірусів зі швидкістю rVZ і гинуть зі швидкістю uI (середня тривалість життя інфікованої клітини $1/u$). З інфікованих клітин виробляються вільні віруси зі швидкістю kI і спадає зі швидкістю vV (середній час життя вільної вірусної частинки дорівнює $1/v$). Швидкість проліферації CTL у відповідь на антиген визначається за допомогою aIT .

За відсутності стимуляції CTL розпадаються зі швидкістю nT . Інфіковані клітини вбиваються CTL зі швидкістю sIT . Внутрішньоклітинна затримка, τ , представляє час, необхідний для інфікованих клітин, щоб виробити віріони після проникнення вірусу [12],

що називається фазою затемнення [25]. Модель задається такою системою звичайних диференціальних рівнянь:

$$\begin{aligned}\frac{dZ(t)}{dt} &= \lambda - mZ(t) - rV(t)Z(t), \\ \frac{dI(t)}{dt} &= rV(t-\tau)Z(t-\tau) - uI(t) - sI(t)T(t), \\ \frac{dV(t)}{dt} &= kI(t) - vV(t), \\ \frac{dT(t)}{dt} &= aI(t)T(t) - nT(t).\end{aligned}\tag{172}$$

Початковими умовами для системи (153) є

$$\begin{aligned}Z(\theta) &= \varphi_1(\theta), I(\theta) = \varphi_2(\theta), V(\theta) = \varphi_3(\theta), T(\theta) = \varphi_4(\theta), \\ -\tau \leq \theta \leq 0, \quad \varphi &= [\varphi_1, \varphi_2, \varphi_3, \varphi_4]^T \in C([-\tau, 0]).\end{aligned}\tag{173}$$

7.6.3 Оптимальне управління ВІЛ-моделі з затримками

У системі людини молекули РНК виробляються з ДНК. Проте, є ферменти, які здійснюють зворотний процес, тобто ДНК можна отримати з РНК. Такий фермент називається зворотною транскриптазою. Один вид таких ферментів є у ВІЛ-1. В результаті, коли заражається частинкою вірусу Т-клітина, вона потрапляє в ядро клітини і виробляє зворотну транскриптазу перетворення вірусних молекул РНК у вірусні молекули ДНК, які потім об'єднуються з молекулами ДНК СТЛ. Отже, СТЛ працюють на створення нових вірусів замість того, щоб виконувати захисну роботу, яку вони мають виконувати в імунітеті системи. У наш час існують препарати, які можуть пригнічувати зворотну транскриптазу, яка дозволяє СТЛ продовжувати свою природну роботу.

Запровадимо функцію керування $c(t)$ у модель (153), $t \in [0, t_f]$, що представляє ефективність інгібіторів зворотної транскриптази, які блокують нову інфекцію. З огляду на важливість фармакологічної затримки у лікуванні ВІЛ розглянемо дискретну затримку в керуючій змінній $c(t)$, яку позначимо ξ , що являє собою затримку, яка виникає між введенням препарату та його появою всередині клітин через час, необхідний для всмоктування, розподілу та проникнення ліків у клітини-мішені. Система керування з дискретним часом затримки у змінних стану та керування:

$$\begin{aligned}\frac{dZ(t)}{dt} &= \lambda - mZ(t) - (1 - c(t - \xi))rV(t)Z(t), \\ \frac{dI(t)}{dt} &= (1 - c(t - \xi)) rV(t - \tau)Z(t - \tau) - uI(t) - sI(t)T(t), \\ \frac{dV(t)}{dt} &= kI(t) - vV(t), \\ \frac{dT(t)}{dt} &= aI(t)T(t) - nT(t).\end{aligned}\tag{174}$$

Початкові умови для змінних стану I і T і, внаслідок затримок, початкові функції для змінних стану Z і V і керування c , задані як

$$I(0) = 3, T(0) = 20, Z(t) = 45, V(t) = 75, -\tau \leq t \leq 0, c(t) = 0, -\xi \leq t < 0.\tag{175}$$

Функція керування $c(t)$ обмежена між 0 і 1. Якщо вона приймає значення 0, медикаментозна терапія для реверсії транскриптази не ефективна. Якщо управління приймає значення 1, воно буде 100% ефективним. Точніше, розглядаємо наступний набір допустимих функцій керування:

$$\Theta = \{c(\cdot) \in L^1([0, t_f], \mathbf{R}) \mid 0 \leq c(t) \leq 1, \forall t \in [0, t_f]\}. \quad (176)$$

Розглянемо цільовий функціонал L^1

$$J[c(\cdot)] = \int_0^{t_f} [V(t) + w \cdot c(t)] dt, \quad (w \geq 1). \quad (177)$$

який вимірює концентрацію вірусу та витрати на лікування за період $[0, t_f]$.

Задача оптимального керування полягає у визначенні функції управління $c(\cdot) \in L^1([0, t_f], \mathbf{R})$, яка мінімізує вартість функціоналу (158) підлягає система керування (155), початкові умови (156) та керуючі обмеження (157). У розділі 4.2 представлені чисельні розв'язки для трьох випадків затримок τ і ξ і ваги $w=1$, $w=5$. Щоб застосувати умови оптимальності, запровадимо змінні стану із затримкою $\xi(t) = Z(t - \tau)$, $\eta(t) = V(t - \tau)$ та змінну управління $\omega(t) = c(t - \xi)$. Використовуючи сполучену змінну $\lambda = (\lambda_Z, \lambda_I, \lambda_V, \lambda_T) \in \mathbf{R}^4$, Гамільтоніан для функціоналу вартості (158) і системи керування (155) задається як

$$\begin{aligned} H(Z, \xi, I, V, T, \eta, c, \omega) = & V + wc + \lambda_Z(\lambda - mZ - (1 - \omega)rVZ) \\ & + \lambda_I[(1 - \omega)r\eta\xi - uI - sIT] + \lambda_V(kI - vT) + \lambda_T(aIT - nT). \end{aligned}$$

Приєднані рівняння задаються

$$\begin{cases} \frac{d\lambda_Z(t)}{dt} = -H_Z[t] - \chi_{[0, t_f - \tau]} H_\xi[t + \tau], \\ \frac{d\lambda_V(t)}{dt} = -H_V[t] - \chi_{[0, t_f - \tau]} H_\eta[t + \tau], \\ \frac{d\lambda_I(t)}{dt} = -H_I[t], \\ \frac{d\lambda_T(t)}{dt} = -H_T[t]. \end{cases} \quad (178)$$

де індекси біля H позначають частинні похідні, а $\chi_{[0, t_f - \tau]}$ – характеристична функція в інтервалі $[0, t_f - \tau]$.

Оскільки термінальний стан вільний, тобто $[Z(t_f), V(t_f), I(t_f), T(t_f)] \in \mathbf{R}^4$ умови трансверсальності є

$$\lambda_Z(t_f) = \lambda_V(t_f) = \lambda_I(t_f) = \lambda_T(t_f) = 0.$$

Для характеристики оптимального керування c введемо таку функцію перемикання:

$$\begin{aligned} \phi(t) = & H_c[t] + \chi_{[0, t_f - \xi]}(t) H_\omega + \xi] \\ = & \begin{cases} 1 + \lambda_Z(t + \xi)rV(t + \xi)Z(t + \xi) - \lambda_I(t + \xi)r\eta(t + \xi)\xi(t + \xi) & 0 \leq t \leq t_f - \xi, \\ 1 & t_f - \xi \leq t \leq t_f. \end{cases} \end{aligned} \quad (179)$$

Умова мінімальності принципу мінімуму [10] дає закон управління

$$c(t) = \begin{cases} 1 & \phi(t) < 0, \\ 0 & \phi(t) > 0, \end{cases} \quad (180)$$

Подібні аргументи також можна використовувати для вирішення пов'язаних проблем оптимального керування, наприклад, можна розглянути додаткове обмеження на кінцеву концентрацію вірусу або включення кінцевих значень цієї концентрації до функціоналу вартості.

Чисельне моделювання

У цьому розділі ми чисельно вивчаємо розв'язок задачі оптимального керування. Розглянемо початкові умови (161) і значення параметрів, наведені в таблиці

Ошибка! Источник ссылки не найден., які базуються на [12].

Табл.7.4 Параметри моделі

Параметр	Значення	Розмірність
λ	5	1/(доба·mm ³)
m	0,03	1/доба
r	0,0014	mm ³ /(virion· доба)
u	0,32	1/доба
s	0,05	mm ³ /доба
k	153,6	1/доба
v	1	1/доба
a	0,2	1/доба
n	0,3	1/доба
t_f	50	доба
τ	0,5	доба
ξ	0,2	доба

Задача оптимального керування із затримками стану та керування

У цьому розділі представлені чисельні розв'язки задачі оптимального керування із затримкою (155)--(158) в інтервалі часу [0; 50] днів і розглянемо три випадки:

Випадок 1: $\tau = \xi = 0$ (без затримок);

Випадок 2: $\tau = 0,5$; $\xi = 0$ (лише внутрішньоклітинна затримка);

Випадок 3: $\tau = 0,5$; $\xi = 0,2$ (внутрішньоклітинна затримка τ і фармакологічна затримка ξ).

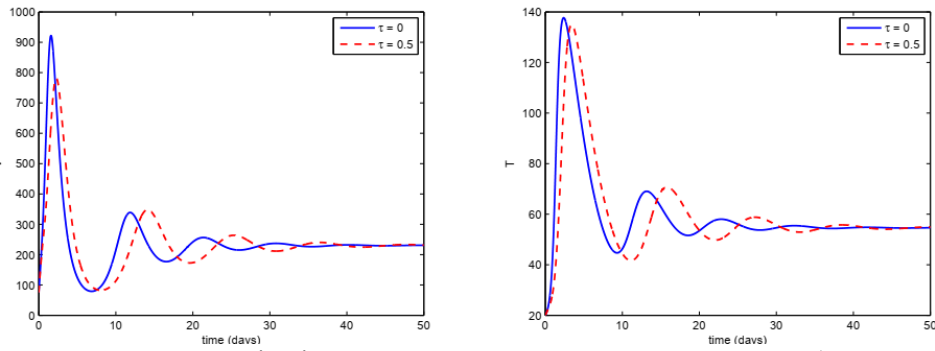


Рис. 7.5. Змінні стану з часовою затримкою $\tau = 0,5$ (пунктирні криві) та без затримки (неперервні криві)

Розглянемо значення параметрів з табл. 7.4 та вагові параметри $w=1$ і $w=5$ у функціоналі вартості (158). Для вирішення задачі оптимального управління (155)–(158) ми дискретизуємо задачу керування та вирішуємо нелінійну оптимізаційну задачу (NLP). У всіх трьох випадках обчислень елементи керування бувають неперевіреними тільки один перемикач на t_s :

$$c(t) = \begin{cases} 1 & 0 \leq t \leq t_s, \\ 0 & t_s \leq t \leq 50. \end{cases} \quad (181)$$

Для ваги $w=1$ отримуємо такі числові результати:

Випадок 1:

$$J(c) = 475,19, t_s = 47,08, Z(50) = 139,48; \\ I(50) = 0,0365; V(50) = 0,962; T(50) = 0,97 \cdot 10^{-5}.$$

Випадок 2:

$$J(c) = 473,05, t_s = 44,78, Z(50) = 139,45; \\ I(50) = 0,02; V(50) = 0,908; T(50) = 0,97 \cdot 10^{-5}.$$

Випадок 3:

$$J(c) = 556,70; t_s = 44,50; Z(50) = 139,56; \\ I(50) = 0,0192; V(50) = 0,873; T(50) = 0,108 \cdot 10^{-4}.$$

Наближення елементів керування та функцій перемикання поблизу часу перемикання t_s , показано на малюнку 7.6.

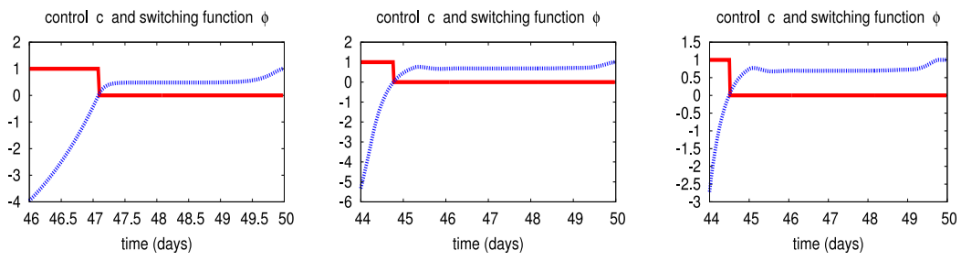


Рис.7.6. Управління Bang-bang $c(t)$ (162) (неперервна крива) і функція перемикання ϕ (160) (пунктирна крива), що відповідає закону керування (161)

Трасекторії станів у трьох випадках дуже схожі на термінальний інтервал $[15; 50]$, тоді як концентрація неінфікованих клітин $Z(t)$ майже ідентична на всьому інтервалі часу

[0; 50]. Щоб відобразити вплив затримок на стан змінних I, V, T , на рис. 7.7 показано порівняння траєкторій стану у випадку 1 (без затримок) і випадок 3 (затримки стану та керування). Ми бачимо, що затримка в s передбачає збільшення концентрації інфікованих клітин $I(t)$ у перші два дні (затримка дії препарату), що також відповідає за збільшення концентрації вільних вірусних частинок V і клітин CTL T .

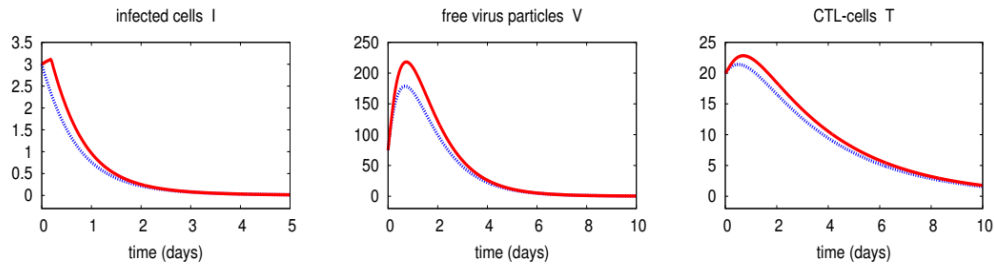


Рис. 7.7. Порівняння траєкторій стану у випадку 1 (без затримок, пунктирна лінія) і випадок 3 (затримки $\tau = 0,5$ і $\xi = 0,2$, неперервні лінії). (зліва) масштабування інфікованих клітин $I(t)$ на $[0; 5]$, (посередині) масштаб вільні вірусні частинки $V(t)$ у $[0; 10]$, (праворуч) масштабування осередків CTL $T(t)$ в $[0; 10]$

7.6.4 Оптимальне управління моделі ВІЛ з лінійною функцією росту антитіл

Математична модель пропонується у такому вигляді:

$$\begin{cases} \frac{dx}{dt} = \lambda - dx - \beta xv, \\ \frac{dy}{dt} = \beta xv - ay - pyz, \\ \frac{dv}{dt} = aNy - \mu v - qvw, \\ \frac{dz}{dt} = cxyz - hz, \\ \frac{dw}{dt} = gxvw - \alpha w, \end{cases} \quad (182)$$

із заданими початковими умовами

$$x(0) = x_0, y(0) = y_0, v(0) = v_0, z(0) = z_0, w(0) = w_0. \quad (183)$$

У цій моделі $x(t), y(t), v(t), z(t), w(t)$ позначають концентрації незаражених клітини, інфіковані клітини, вірус ВІЛ, клітини CTL та антитіла в момент часу t відповідно. Здорові $CD4^+$ Т-клітини (x) ростуть зі швидкістю λ , гинуть зі швидкістю d і інфікуються вірусом зі швидкістю βxv . Інфіковані клітини (y) гинуть зі швидкістю a і гинуть у відповідь на CTL зі швидкістю p . Вільні віруси (v) виробляється інфікованими клітинами зі швидкістю aN і гинуть зі швидкістю μ і розпадаються за присутності антитіл зі швидкістю q , де N є кількість вільних вірусів, вироблених кожною активно інфікованою клітиною упродовж свого життя. CTL (z) розширюються у відповідь на вірусний антиген, отриманий із інфікованих клітин, із швидкістю c і розпад за відсутності антигенної стимуляції зі швидкістю h . Нарешті, антитіла (w) розвиваються у відповідь на

вільний вірус зі швидкістю g і розпадаються зі швидкістю α . Варто зауважити, що всі показники моделі вважаються невід'ємними.

Оптимальне управління

У цьому розділі ми вивчаємо задачу оптимального керування введення медикаментозної терапії у модель (163) за припущення, що лікування зменшує реплікацію вірусу. Наша мета — знайти стратегію лікування (t) , яка максимізує кількість $CD4^+$ Т-клітин, а також кількість СТЛ та відповідь імунних антитіл, зберігаючи вартість, виміряну з точки зору сили хіміотерапії та поєднання тривалості та інтенсивності, якомога менше.

Проблема оптимізації

Для застосування теорії оптимального керування ми пропонуємо таку систему керування з двома змінними керування:

$$\begin{cases} \frac{dx}{dt} = \lambda - dx - \beta(1-u_1(t))xv, \\ \frac{dy}{dt} = \beta(1-u_1(t))xv - ay - pyz, \\ \frac{dv}{dt} = aN(1-u_2(t))y - \mu v - qvw, \\ \frac{dz}{dt} = cxyz - hz, \\ \frac{dw}{dt} = gxvw - \alpha w, \end{cases} \quad (184)$$

Тут u_1 представляє ефективність медикаментозної терапії в блокуванні нової інфекції, так що рівень інфікування в присутності препарату становить $(1-u_1)$, тоді як u_2 означає ефективність лікарської терапії для інгібування продукції вірусу, наприклад, швидкість утворення віріонів під терапією становить $(1-u_2)$. Наша задача оптимізації полягає в максимізації цільового функціоналу:

$$J(u_1, u_2) = \int_0^{t_f} \left\{ x(t) + z(t) + w(t) - \left[\frac{A_1}{2} u_1^2(t) + \frac{A_2}{2} u_2^2(t) \right] \right\} dt, \quad (185)$$

де t_f — період часу обробки і додатні константи A_1 і A_2 стоять на витрати на введене лікування. Дві функції керування, u_1 і u_2 вважаються обмеженими і інтегрованими за Лебегом. Шукаємо u_1^* і u_2^* такі, що

$$\begin{aligned} J(u_1^*, u_2^*) &= \max\{J(u_1, u_2) : (u_1, u_2) \in U\}, \\ U &= \{u_1(\cdot), u_2(\cdot) : u_i(\cdot) \in L^2[0, t_f]; 0 \leq u_i(t) \leq 1, t \in [0, t_f]\} \end{aligned} \quad (186)$$

Зауважимо, що природно максимізувати кількість СТЛ та імунної відповіді у задачі оптимального керування. Дійсно, клінічно було відзначено, що особи, які підтримують високий рівень СТЛ, залишаються здоровими довше. Тому ми бажаємо максимально збільшити кількість СТЛ, щоб гарантувати, що, якщо вірусне навантаження відновиться, імунна система впорається з цим. Необхідно встановити найкращі лікарські засоби, зводячи негативні наслідки до мінімуму.

Система оптимальності

Принцип мінімуму Понтрягіна забезпечує необхідні умови оптимальності такої задачі оптимального керування [19]. Цей принцип перетворює (165), (166) і (167) у задачу мінімізації гамільтоніана H поточною відносно u_1, u_2 , де

$$H = \frac{A_1}{2} u_1^2(t) + \frac{A_2}{2} u_2^2(t) - x - z - w + \sum_{i=0}^5 \lambda_i f_i,$$

$$\begin{cases} f_1 = \lambda - dx - \beta(1-u_1)xv, \\ f_2 = \beta(1-u_1)xv - ay - pyz, \\ f_3 = aN(1-u_2)y - \mu v - qvw, \\ f_4 = cxyz - hz, \\ f_5 = gxvw - \alpha w, \end{cases}$$

Теорема. Для заданих оптимальних керувань u_1^*, u_2^* і розв'язків x^*, y^*, v^*, z^*, w^* відповідної системи станів (165) існують суміжні змінні $\lambda_1 \dots \lambda_5$, що задовольняють рівнянням

$$\begin{cases} \frac{d\lambda_1(t)}{dt} = 1 + \lambda_1(t)[d + (1-u_1^*(t))\beta v^*(t)] - \lambda_2(t)(1-u_1^*(t))\beta v^*(t) \\ \quad - \lambda_4(t)cy^*(t)z^*(t) - \lambda_5(t)gv^*(t)w^*(t), \\ \frac{d\lambda_2(t)}{dt} = \lambda_2(t)(a + pz^*(t)) - \lambda_3(t)(1-u^*(t))aN - \lambda_4(t)cx^*(t)z^*(t), \\ \frac{d\lambda_3(t)}{dt} = \lambda_1(t)(1-u_1^*(t))\beta x^*(t) - \lambda_2(t)(1-u_1^*(t))\beta x^*(t) \\ \quad + \lambda_3(t)(\mu + qw^*(t)) - \lambda_5(t)gx^*(t)w^*(t), \\ \frac{d\lambda_4(t)}{dt} = 1 + \lambda_2(t)py^*(t) + \lambda_4(t)[h - cx^*(t)y^*(t)], \\ \frac{d\lambda_5(t)}{dt} = 1 + \lambda_3(t)qv^*(t) + \lambda_5(t)[\alpha - gx^*(t)v^*(t)], \end{cases} \quad (187)$$

із умовами трансверсальності

$$\lambda_i(t_f) = 0, i = \overline{1,5}. \quad (188)$$

Оптимальне управління задається як

$$\begin{aligned} u_1^*(t) &= \min \left\{ 1, \max \left[0, \frac{\beta}{A_1} ((\lambda_2(t) - \lambda_1(t)x^*(t)v^*(t))) \right] \right\}, \\ u_2^*(t) &= \min \left\{ 1, \max \left[0, \frac{1}{A_2} \lambda_3(t)aNy^*(t) \right] \right\}. \end{aligned} \quad (189)$$

Система оптимальності складається з системи станів (165), сполученої з приєднаними рівняннями (168), початкові умови (164), умови трансверсальності (169) та характеристики оптимальних елементів керування (170). А саме, якщо підставити вирази u^*, u_2^* в (165), отримаємо таку систему оптимальності:

$$\begin{cases}
\frac{dx^*(t)}{dt} = \lambda - dx^*(t) - \beta(1-u_1^*(t))x^*(t)v^*(t), \\
\frac{dy^*(t)}{dt} = \beta(1-u_1^*(t))x^*(t)v^*(t) - ay^*(t) - py^*(t)z^*(t), \\
\frac{dv^*(t)}{dt} = aN(1-u_2^*(t))y - \mu v^*(t) - qv^*(t)w^*(t), \\
\frac{dz^*(t)}{dt} = cx^*(t)y^*(t)z^*(t) - hz^*(t), \\
\frac{dw^*(t)}{dt} = gx^*(t)v^*(t)w^*(t) - \alpha w^*(t), \\
\frac{d\lambda_1(t)}{dt} = 1 + \lambda_1(t)[d + (1-u_1^*(t))\beta v^*(t)] - \lambda_2(t)(1-u_1^*(t))\beta v^*(t) \\
\quad - \lambda_4(t)cy^*(t)z^*(t) - \lambda_5(t)gv^*(t)w^*(t), \\
\frac{d\lambda_2(t)}{dt} = \lambda_2(t)(a + pz^*(t)) - \lambda_3(t)(1-u_2^*(t))aN - \lambda_4(t)cx^*(t)z^*(t), \\
\frac{d\lambda_3(t)}{dt} = \lambda_1(t)(1-u_1^*(t))\beta x^*(t) - \lambda_2(t)(1-u_1^*(t))\beta x^*(t) \\
\quad + \lambda_3(t)(\mu + qw^*(t)) - \lambda_5(t)gx^*(t)w^*(t), \\
\frac{d\lambda_4(t)}{dt} = 1 + \lambda_2(t)py^*(t) + \lambda_4(t)[h - cx^*(t)y^*(t)], \\
\frac{d\lambda_5(t)}{dt} = 1 + \lambda_3(t)qv^*(t) + \lambda_5(t)[\alpha - gx^*(t)v^*(t)],
\end{cases} \quad (190)$$

де u_1^* , u_2^* визначаються виразами (170).

Чисельне моделювання

Для розв'язання системи оптимальності (170),(171) скористаємося чисельною схемою, що заснована на прямих і зворотних кінцевих різницевих апроксимаціях.

Точніше, ми реалізували Алгоритм 1.

Для нашого чисельного моделювання ми обрали наступні параметри (див. табл. 7.5):

$$\begin{aligned}
\lambda &= 1, d = 0,1, \beta = 0,00025, p = 0,001, a = 0,2, c = 0,03, \\
p &= 0,2, g = 0,0001, \alpha = 0,12, q = 0,01, A_1 = 250, A_2 = 2500. \\
\lambda_i(t_f) &= 0, i = \overline{1,5}; \quad u_1(0) = 0, u_2(0) = 0.
\end{aligned}$$

Ці параметри показують стабільність останньої ендемічної точки E_4 з усіма ненульовими компонентами системи. Початкові значення кожного компонента системи задається таким чином:

$$x_0 = 5, y_0 = 1, z_0 = 2, w_0 = 1, v_0 = 1.$$

Табл.7.5. Параметри моделі

Пар.	Значення	Значення
λ	вихідний показник CD4+ Т-клітин	1--10 кліт/(мкл доба)
d	швидкість розпаду здорових клітин	0,007--0,1 1/доба
β	швидкість, з якою CD4+ Т-кліт. інф.	0,00025--0,5
		мкл/(віріон доба)
a	смертність від інф. CD4+ Т клітин	0,2--0,3 1/доба
μ	швидкість очищення вірусу	2,06--3,81 1/day
N	кількість віріонів	6,25--23600 1/віріон
ρ	швидкість виведення інфекції	$(0,1 \div 0,405)10^{-3}$
		мл віріон/доба
c	швидкість активації клітин CTL	0,0051--3,912 1/day
h	смертність клітин CTL	0,004--8,087 1/day
q	Швидкість нейтралізації віріонів	0,12 1/day
g	швидкість активації антитіл	0,00013 1/day
α	Смертність антитіл	0,12 1/day

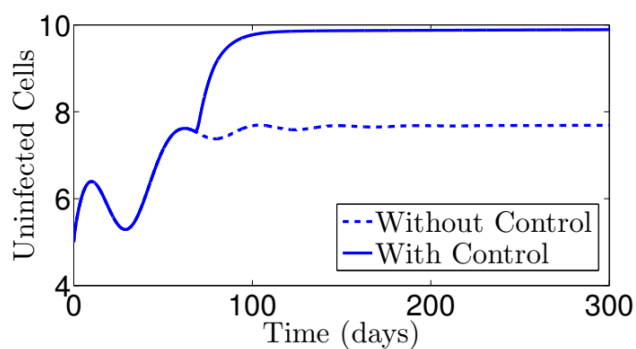


Рис.7.7. Еволюція неінфікованих клітин у часі

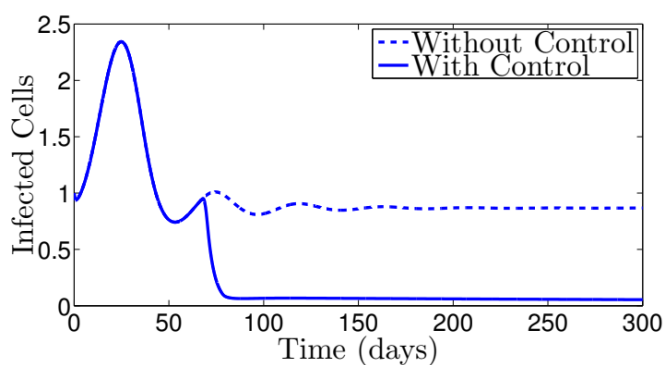


Рис. 7.8. Еволюція інфікованих клітин у часі

На рис. 7.8 добре видно, що після введення терапії популяція неінфікованих клітин значно зростає порівняно із безконтрольними. На рис. 7.9 показано, що з управлінням кількість інфікованих клітин значно зменшується після кількох тижнів терапії. На рис. 7.10 показано, що з управлінням вірусне навантаження знижується до дуже низького рівня після перших днів терапії, тоді як без управління вона залишається значно вищою. Це свідчить про вплив призначеної терапії для реплікації вірусу. На рис. 7.11 показано

адаптивну імунну відповідь як функцію часу. Адаптивний імунітет чітко виражений під впливом управління.

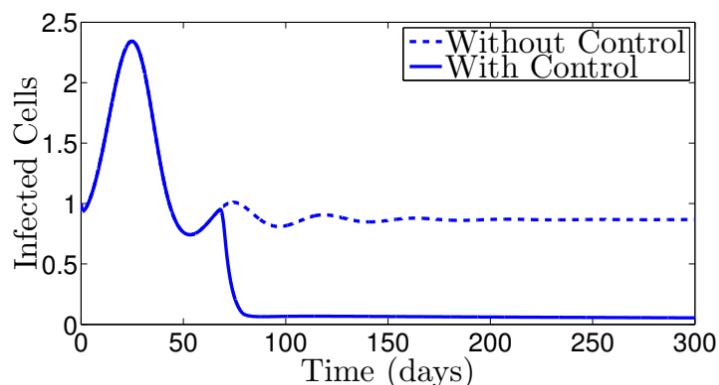


Рис. 7.9.Еволюція вірусу ВІЛ у часі

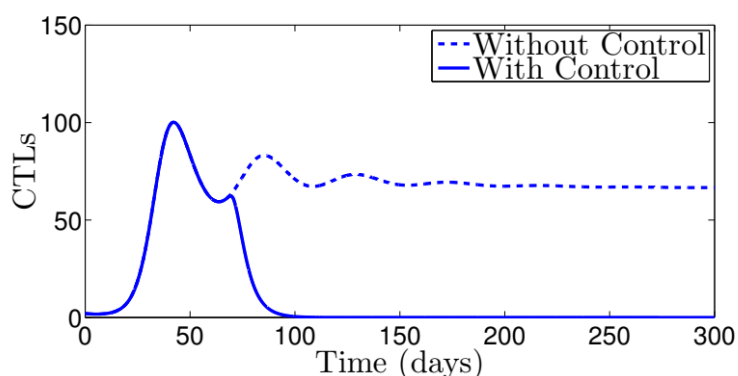


Рис. 7.10.Еволюція СТЛ-клітин у часі

Відзначимо, що всі криві (без управління) попередніх фігур сходяться до ендемічної точки з координатами (7,6923; 0,8666;120; 66,2721; 48,888). Цей результат добре узгоджується з результатом пропозиції 6, оскільки з обраними нами параметрами ми маємо $R_0 = 2,0833 > 1$; $R_1^{CTL,W} = 1,2037 > 1$; $R_2^{CTL,W} = 1,3313 > 1$.

Бібліографія

1. Асаченков А.Л., Марчук Г.И. Уточнённая математическая модель инфекционного заболевания *Математическое моделирование в иммунологии и медицине/под ред. Г.И. Марчука*. Новосибирск: Наука, 1982, С. 44--58.
2. Болодурина И.П., Луговская Ю.П. Оптимальное управление иммунологическими реакциями организма человека *Проблемы управления*, 2009, № 5, С. 44--52.
3. Бочаров Г.А. Численные эксперименты с математической моделью Т-клеточного иммунного ответа *Вычислительные процессы и системы*, Вып. 3 / под ред. Г.И. Марчука. М.: Наука, 1985, С. 208--217.
4. Бочаров Г.А., Марчук Г.И. Прикладные проблемы математического моделирования в иммунологии *Журнал вычислительной математики и математической физики*, 2000, Т. 40, № 12, С. 1905--1920
5. Зеленський К., Бовсуновська К., Болховітін В. Алгоритмічне забезпечення

розв'язання нелінійних крайових задач тепломасопереносу. *Modern engineering and innovative technologies*, Issue 15, Part 1, 2021, p. 6--11.

6. Инфекционные болезни человека. Динамика и контроль (пер. с англ. А. А. Романюхи, С. Г. Руднева) / под ред. Г. И. Марчука. М. : Научный мир, 2004, 784 с.

7. Марчук Г.И. Математические модели в иммунологии. *Вычислительные методы и эксперименты*, М.: Наука, 1991, 304 с.

7. Марчук Г.И., Романюха А.А., Бочаров Г.А. Математическое моделирование противовирусного иммунного ответа при вирусном гепатите В. *Математические вопросы кибернетики*. Вып. 2 / под ред. С.В. Яблонского. – М.: Наука, 1989, С. 5--70.

8. Романюха А.А., Руднев С.Г., Зуев С.М. Анализ данных и моделирование инфекционных заболеваний *Современные проблемы вычислительной математики и математического моделирования*. Т. 2. Математическое моделирование / под ред. В.П. Дымникова. – М.: Наука, 2005, С. 352--404.

9. Трусов П.В. и др. Регуляция противовирусного иммунного ответа организма: математическая модель, качественный анализ, результаты *Мат. биол. и биоинф.*, 2018, №13(2), С. 402--425.

10. Ешли Н.Т., Демас Г.Е. Нейроендокринно-імунні схеми, фенотипи та взаємодії. *Гормони і поведінка*, 2017, т. 87, с. 25--34.

11. Banks H., Kabanikhin S. Numerical algorithm for constructing an individual mathematical model of HIV dynamics at cellular level *J. Inverse Ill-Posed Probl.*, 2018, V. 26, № 6, P. 859--873.

13. Bellavans M., Rivest S. Neuroendocrine control of innate immune system in health and brain disease. *Immunological reviews*, 2012, vol. 248, vol. 1, p. 36--55.

14. Forys U. Global analysis of Marchuk's model in a case of strong immune system *J. Biol. Sys.*, 2000, V. 8, N 4, P. 331--346.

15. Graw F., Perelson A. S. Modeling Viral Spread *Annual review of virology*. 2016, Vol. 3, P. 555--572.

16. Kabanikhin S.I., Krivorotko O.I. Inverse problems of immunology and epidemiology *Eurasian Journal of Math. and Computer Applications*. 2017, V. 5, Issue 2, P. 14--35.

17. Marchuk G.I. Mathematical modeling of immune response in infectious diseases. – Dordrecht: Springer Science & Business Media, 2013, 350 p.

17. Marchuk G., Petrov R., Romanyukha A., Bocharov G. Mathematical model of antiviral immune response. I. Data analysis, generalized picture construction and parameters evaluation for hepatitis B. *Journal of Theoretical Biology*, 1991, vol. 151, no. 1, pp. 1--40.

21. Belda O., Targett-Adams P. Small molecule inhibitors of the hepatitis C virus-encoded NS5A protein. *Vir. Res.*, 2012, 170, p. 1--14.

22. Knodel M., Reiter S., Targett-Adams P. 3D Spatially Resolved Models of the Intracellular Dynamics of the Hepatitis C Genome Replication Cycle. *Viruses*, 2017, 9, 282.

23. Knodel M., Nagel A., Reiter S., Rupp M. Quantitative analysis of Hepatitis C NS5A viral protein dynamics on the ER surface. *Viruses*, 2018, 10, 27.

24. Knodel M., Nagel A., Reiter S., Rupp M. Multigrid analysis of spatially resolved hepatitis C virus protein simulations. *Comp. Vis. Sci.*, 2015, 17, p. 235--253.

25. Knodel M., Nagel A., Reiter S., Rupp M. On Estimation of a Viral Protein Diffusion Constant on the Curved Intracellular ER Surface. Springer International Publishing: Cham, Switzerland, 2016; pp. 641--657.

26. Knodel M., Targett-Adams P., Grillo A.. Advanced Hepatitis C Virus Replication PDE Models within a Realistic Intracellular Geometric Environment. *International Journal of Environmental Research and Public Health*, 2019, 16, 53 p.

27. Kohli A., Shaffer, A., Sherman, A. Treatment of hepatitis C: A systematic review.

JAMA, 2014, 312, p. 631–640.

27. Li J., Kuang Y. Analysis of a model of the glucose-insulin regulatory system with two delays. *SIAM J Appl Math.*, 2007, 67, pp. 757–76.

28. Lindenbach B., Rice C. The ins and outs of hepatitis C virus entry and assembly. *Nat. Rev. Microbiol.*, 2013, 11, p. 688–700.

30. Martin N., Pitcher A., Vickerman P.. Optimal Control of Hepatitis C Antiviral Treatment. Programme Delivery for Prevention amongst a Population of Injecting Drug Users. *PLoS One*, 2011, V.6, No 8, e23309, 17 p.

31. Pachpute G., Chakrabarty S. Dynamics of hepatitis C under optimal therapy and sampling based analysis. *Commun Nonlinear Sci. Numer. Simul.*, 18, 2013, pp. 2202–11.

32. Rong L., Guedj J., Dahari H. Analysis of hepatitis C virus decline during treatment with the protease inhibitor danoprevir using a multiscale model. *PLoS Comput. Biol.*, 2013, 9, e1002958.

41. Allali K., Harroudi S. Optimal control of an hiv model with a trilinear antibody growth function. *Discrete and continuous dynamical systems series S*, V. 15, N. 3, 2022, pp. 501–517.

42. Allali K., Harroudi S., Torres D. Analysis and optimal control of an intracellular delayed HIV model with CTL immune response, *Math. Comput. Sci.*, 12, 2018, pp. 111–127.

43. Allali K., Tabit Y., Harroudi S. On HIV model with adaptive immune response, two saturated rates and therapy, *Math. Model. Nat. Phenom.*, 12, 2017, pp. 1–14.

44. Ciupe M., Bivort B., Bortz D. Estimating kinetic parameters from HIV primary infection data through the eyes of three different mathematical models, *Math. Biosci.*, 200, 2006, pp. 1–27.

45. Culshaw R., Ruan S., Spiteri R. Optimal HIV treatment by maximising immune response, *J. Math. Biol.*, 48, 2004, pp. 545–562.

46. Culshaw R., Ruan S., Webb G. A mathematical model of cell-to-cell spread of HIV-1 that includes a time delay. *J Math Biol.*, 2003, 46, pp. 425–44.

47. Miao H., Xia X., Wu H. On Identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev. Soc. Ind. Appl. Math.*, 2011, V. 53, Issue 1, P. 3–38.

47. Nelson P., Perelson A. Mathematical analysis of delay differential equation models of HIV-1 infection. *Math. Biosci.*, 2002, No. 179, P. 73–94.

48. Pawelek K., Liu S., Pahlevani F., Rong L. A model of HIV-1 infection with two time delays: mathematical analysis and comparison with patient data. *Math. Biosci.*, 2012, V. 235, No. 1, P. 98–108.

50. Pitchaimani M., Monica C. Global stability analysis of HIV-1 infection model with three time delays. *J. Appl. Math. Comput.*, 2015, V. 48, P. 293–318.

51. Wang J., Lang J., Zou X. Analysis of an age structured HIV infection model with virus-to-cell infection and cell-to-cell transmission. *Nonlinear Analysis: Real World Applications*, 2017, V. 34, P. 75–96.

52. Winter F., Andersen M., Ottesen J. The minimal model of the hypothalamic-pituitary-adrenal axis. *Journal of Mathematical Biology*, 2011, vol. 63, no. 4, pp. 663–690.

53. Stafford M., Corey L, Cao Y. Modeling plasma virus concentration during primary HIV infection. *J Theor Biol.*, 2000, 203, pp. 285–301.

54. Wang L, Li M. Mathematical analysis of the global dynamics of a model for HIV infection of CD4+ T cells. *Math Biosci.*, 200, 2006, pp. 44–57.

55. Zhu H, Zou X. Dynamics of a HIV-1 infection model with cell-mediated immune response and intracellular delay. *Disc Cont Dyan Syst B*. 12, 2009, pp. 511–24.

Розділ 8 Моделі ракових пухлин

8.1 Загальні засади лікування ракових пухлин

Необхідність узагальнення наявних даних про патогенез раку спонукають до створення математичних моделей, що описують причинно-наслідкові зв'язки, які визначають динаміку процесу пухлинного росту. З практичної точки зору використання методу математичного моделювання в онкології надає можливість для кількісного та часового прогнозування протікання хвороби і впливу різних видів лікування на організм пацієнта на різних рівнях: клітин, тканин, організму у цілому. Оскільки ріст пухлини здійснюється швидкими темпами, в пухлинних тканинах не встигає сформуватися система кровозабезпечення, що забезпечує доставку кисню до клітин. Це призводить до утворення ділянок некрозу, тобто накопичення пухлинних клітин, що загинули внаслідок нестачі кисню та поживних речовин.

Генетична неоднорідність пухлинних клітин, що спричиняється високим рівнем генетичних мутацій в пухлинній популяції у процесі розвитку онкологічного захворювання, є значна проблема, оскільки призводить до резистентності пухлин до різних видів лікування.

Хіміотерапевтичне лікування є основним методом терапії в онкології, тому низка праць присвячена моделюванню впливу цього виду лікування на ріст пухлин.

Імунологічне лікування, спрямоване на стимуляцію реакції імунної системи на ріст пухлин, є не настільки широкоживаним на практиці методом, як хіміотерапія і променева терапія. Але він приваблює тим, що він суттєво менш токсичний по відношенню до нормальних тканин.

Перспективним спрямуванням у терапії раку є вірусотерапія, тобто використання для лікування онкологічних захворювань вірусів, які уражають пухлинні клітини. Існує кілька механізмів ураження ракових клітин онкологічними вірусами:

- багатократна реплікація вірусу всередині ракової клітинки, що призводить її до загибелі;
- вироблення цитотоксичного протеїну, що пошкоджує ракові клітини;
- стимулювання імунної системи організму у зв'язку із процесом, індукованим вірусом.

Онкологічне захворювання являє собою появу в організмі та розмноження клону або кількох клонів клітин, проліферативна активність яких не підпорядковується регуляторними механізмам, припадними нормальним клітковим популяціям. Життєвий цикл клітини складається із таких основних етапів:

Стадія росту (G_1). У цю стадію клітини опиняються одразу після ділення материнської клітини на дві дочірні. Здійснюється збільшення цитоплазми, синтез та накопичення речовин, що необхідні для подвоєння ДНК. Клітина у цій стадії перебуває фіксований час.

Стадія спокою (G_0). У цій стадії клітина опиняється на стадії росту. Якщо у популяції клітин спостерігається нестача зрілих клітин, клітини у стадії G_0 можуть переходити у стадії ділення і зрілості. Пухлинні клітини здатні затримуватися у стадії спокою, а потім знову повертатися у стадію ділення.

Стадія ділення ($S - G_2 - M$). У цій стадії опиняються клітини зі стадії спокою за нестачі зрілих клітин. Стадія ділення поділяється на три фази.

- S -- синтетичний період, що характеризується подвоєнням молекул ДНК;

- G_2 -- постсинтетичний період, в якому здійснюється інтенсивний синтез білків;
- M -- мітоз, в якому хромосоми розщеплюються, ядро роздвоюється, цитоплазма ділиться навпіл і в результаті утворюються дві клітини.

Стадія зрілості (диференцировки) (Z). У цій стадії клітини опиняються зі стадії спокою за нестачі диференцированих клітин у популяції. Після попадання у цей стан клітинка трансформується у диференцировану, тобто здатну виконувати специфічну функцію в організмі, властиву даному виду клітин. Виконуючи свою функцію в організмі, вона зношується і гине.

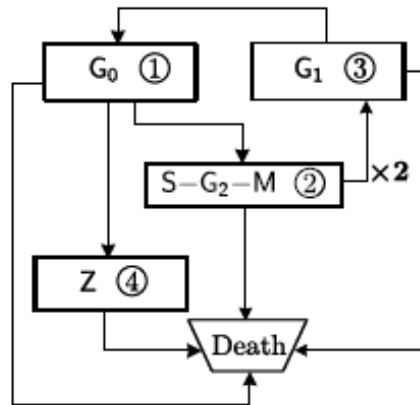


Рис.8.1 Граф переходів клітинки між стадіями життєвого циклу

Сучасні протипухлинні препарати вибірково впливають на визначені структури клітин у різних етапах життєвого циклу: антиметаболіти і дактиноміцин діють на клітини у S -фазі, блеоміцин ефективний у G_2 -фазі, винкристин діє в M -фазі. Важливо, що усі клітинки у стадії спокою, у тому числі й пухлинні, нечутливі до дії протипухлинних хіміопрепаратів та до променевого впливу. Такі клітинки називаються клоногенними; вони можуть бути джерелом рецидивів пухлин.

Хіміотерапія - це лікування, яке призначається для уповільнення або зупинки розвитку раку вбиваючи ракові клітини, і його дають циклами. Цикл — це час між двома введеннями препарату; цей час необхідний для відновлення організму від побічної дії препарату. Типовий цикл хіміотерапії становить 21 або 28 днів. У клінічних дослідженнях метастатичного раку молочної залози з комбінацією хіміотерапевтичного препарату та анти-VEGF два препарати вводять одночасно у кожному циклі. Оскільки анти-VEGF знижує перфузію хіміотерапевтичних препаратів [I1], виникає питання, чи не краще вводити два препарати роздільно у часі. Нещодавні експерименти на мишах показали, що у такому випадку справді призводить до зменшення об'єму пухлини [I2].

У випадку метастатичного раку молочної залози були проведені клінічні випробування з комбінацією хіміотерапевтичного препарату та анти-VEGF, де вводять ці два препарати одночасно в кожному циклі. Але анти-VEGF негативно впливає на ефективність хіміотерапії, оскільки знижує перфузію хіміотерапевтичних препаратів. Отже, виникає питання, чи було б корисніше вводити обидва препарати не одночасно.

Розробляється математична модель комбінованої терапії раку за допомогою анти-VEGF (наприклад, бевацизумаб, сунітініб) і хіміотерапії (наприклад, доцетаксел, паклітаксел) і використовується модель для порівняння двох схем: одночасних ін'єкцій і послідовних ін'єкцій (не перекриваються). Модель представлено системою диференціальних рівнянь у частинних похідних (PDE) з такими змінними: густина ракових

клітин, дендритних клітин, $CD8^+$ Т-клітин та ендотеліальних клітин і концентрації IL-12, VEGF, кисню та двох препаратів. Модель заснована на мережі, показаній на рис. 7.2.

Крім того, імуногенність раку активує дендритні клітини, які потім активують $CD8^+$ Т-клітини за допомогою IL-12, і активовані Т-клітини вбивають ракові клітини. VEGF секретується раковими клітинами, притягує ендотеліальні клітини в пухлину, ініціюючи процес ангіогенезу, який приносить кисень до ракових клітин. VEGF також блокує дозрівання дендритних клітин і $CD8^+$ Т-клітин. На рис. 7.2 показано також вплив два препарати.

Ефективність лікування вимірюється за кількістю днів, необхідних для зменшення початкового об'єму пухлини на 98%.

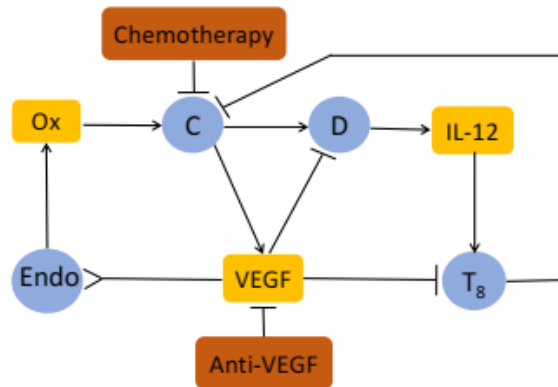


Рис.8.2. Взаємодія імунних клітин із раковими

Гострі стрілки вказують на поширення/активацію, заблоковані стрілки вказують на вбивство/блокування, а перевернута стрілка вказує вербування/ хіміоатракціон.

Метою імунотерапії є посилення природної здатності організму боротися з раком підвищення ефективності імунної системи, використання звичайної хіміотерапії раку в поєднанні з імунотерапією вважалось недоцільним, оскільки хіміотерапія взагалі знижує імунітет і може звести нанівець переваги імунотерапії коли дають разом, хоча теоретичні дослідження та математичні праці показали, що поєднання хіміотерапії з імунотерапією може мати додаткові ефекти, що підвищує ефективність лікування раку. Логіка розробки стратегії комбінованої хіміоімунотерапії базується на використанні мінімальної кількості хіміотерапевтичних препаратів для ефективного знищення пухлинних клітин і застосування імунотерапії для підтримки імунної системи пацієнта, й отже, зміцнення природних захисних сил організму проти як пухлинні клітини, так і небезпечних побічних ефектів хіміотерапії.

Коли рак вражає клітини людини, перша лінія захисту людського організму є природний вбивця (NK) клітини: білі кров'яні клітини, які активно сканують тіло для аномальних клітин, як тільки знайдено, бойові групи і виживає найсильніший. Як частина специфічної імунної відповіді втручаються $CD8^+$ цитотоксичні Т-лімфоцити (CTL), щоб убити пухлинні клітини. У цій роботі ми використовуємо існуючу модель, [1.2], яка описує взаємодію пухлинних клітин, NK-клітин, $CD8^+$ Т-клітин, і циркулюючі клітини лімфоцитів при комбінованій імунотерапії та хіміотерапії, потім застосовуємо метод теорії оптимального керування з квадратичним керуванням.

Існує кілька різних типів імунотерапії, для яких використовуються контрастні методи для боротьби з пухлинними клітинами. Вони поділяються на три основні категорії: модифікатори імунної відповіді, моноклональні антитіла та вакцини. Перша категорія містить речовини, які впливають на імунну відповідь, такі як інтерлейкіни (включаючи IL-

2), інтерферони, фактори некрозу пухлин (TNF), колонієстимулюючі фактори (CSF), і фактори росту В-клітин. У наступній категорії у даний час розробляються моноклональні антитіла, спрямовані на специфічні ракові антигени. Третя категорія - це вакцини, які загалом використовуються в терапевтичних цілях і створюються з пухлинних клітин.

8.2 Оптимальне управління імунотерапією пухлин

8.2.1 Математична модель

Розробка моделі

Мета створення математичної моделі полягає у забезпеченні достатньої складності, щоб модель якісно генерувала клінічно спостережувану *in vivo* модель зростання пухлини, водночас зберігаючи достатню простоту для аналізу. Модель, яка розглядається нижче, є система звичайних диференціальних рівнянь, змінними стану якої є популяції пухлинних клітин, специфічних та неспецифічних імунних клітин і концентрації терапевтичних втручань.

Модель, яку ми тут пропонуємо, має спільні риси з моделлю, [p19]. Модель [p18] містить два важливі уточнення: автори розрізняють між популяціями та діями специфічних (таких як $CD8^+$) і відстежувати їх неспецифічні (такі як NK) ефекторні клітини та дозволяють вбудовувати додаткові імунотерапія у вигляді ін'єкцій інтерлейкіну-2 (IL-2) і лімфоцитів, що інфільтрують пухлину (TIL).

Модельні припущення

Слід зазначити, що немає універсальної згоди щодо основної динаміки або точного каскаду подій, які відбуваються у процесі імунної відповіді. Припущення також базуються на опублікованих заявах.

- Пухлина росте логістично за відсутності імунної відповіді.
- NK (природні кілери, неспецифічні ефектори), і $CD8^+$ (цитотоксичні Т-лімфоцити), специфічні ефекторні клітини, здатні вбивати пухлинні клітини.
- NK і $CD8^+$ Т-клітини реагують на пухлинні клітини розширенням і збільшенням цитолітичної активності.
- NK-клітини зазвичай присутні в організмі, навіть якщо немає пухлинних клітин, оскільки вони є частиною вродженої імунної відповіді.
- У рамках специфічної імунної відповіді активні пухлиноспецифічні $CD8^+$ Т-клітини присутні лише у великій кількості, коли присутні пухлинні клітини.
- NK і $CD8^+$ Т-клітини стають неактивними після деякої кількості зустрічей з пухлинними клітинами.

Наступні додаткові припущення використовуються при розробці терапевтичних умов.

- Рівень циркулюючих лімфоцитів можна використовувати як міру здоров'я пацієнта. Зустрічі між циркулюючими лімфоцитами та пухлинними клітинами збільшують популяцію ефекторних клітин.
- Частка пухлинної популяції, знищеної хіміотерапією, залежить від кількості препарату у системі.

- Під дією хіміотерапії також гинуть: частка NK-клітин, $CD8^+$ Т-клітин та циркулюючих лімфоцитів.
- Т-клітини є компонентами процесу стимуляції та елімінації активованих ефекторних клітин, спрощення моделі призначене для відображення саморегуляційної природи імунної системи.

Модель ОДУ: загальний вигляд рівнянь

Модель описує кінетику чотирьох популяцій (пухлинних клітин і трьох типів імунних клітин), а також дві концентрації препарату у крові, використовуючи низку пов'язаних звичайних диференціальних рівнянь.

- $T(t)$ -- популяція пухлинних клітин;
- $L(t)$ -- пухлиноспецифічна популяція $CD8^+$ Т-клітин;
- $N(t)$ -- популяція неспецифічних ефекторно-імунних клітин;
- $C(t)$ -- популяція циркулюючих лімфоцитів;
- $I(t)$ -- концентрація ІЛ-2 (МО на літр);
- $M(t)$ -- концентрація хіміопрепарату (мг на літр).

Рівняння, що визначають кінетику популяції, мають враховувати члени зростання для кожної популяції, часткове знищення клітин, рекрутування клітин, клітинна інактивація та зовнішнє медикаментозне втручання. Ми намагаємося використовувати найпростіші вирази для кожного члена, які ще якісно відображають експериментальні дані і розпізнають взаємодію популяції.

Розглядаються фактичні рівні популяції NK і $CD8^+$ Т-клітин, щоб представити «ефективність» імунної клітини у тому сенсі, що підрахунок збільшення загальної кількості популяції у математичній моделі може підтверджуватися також біологічно як більша кількість загальних клітин, або як кожна окрема імунна клітина ефективніше вбиває цільові пухлинні клітини.

Система рівнянь, що описує ріст, смерть і взаємодію цих популяцій з лікуванням має вигляд

$$\frac{dT}{dt} = aT(1-bT) - cNT - DT - K_T T, \quad (191)$$

$$\frac{dN}{dt} = \alpha_1 - fN + g \frac{T^\eta}{h + T^\eta} N - pNT - K_N MN, \quad (192)$$

$$\frac{dL}{dt} = -mL - uL^2 - qLT + \frac{p_I I}{g_I + I} L + V_L(t) - K_L ML, \quad (193)$$

$$\frac{dC}{dt} = \alpha_2 - \beta C - K_C MC, \quad (194)$$

$$\frac{dM}{dt} = -\gamma M + v_M(t), \quad (195)$$

$$\frac{dI}{dt} = -\mu_I I + v_I(t) + wLI - \mu_I I, \quad (196)$$

де $D = d \frac{(L/T)^l}{s + (L/T)^l}$, $T(0) = T_0$, $N(0) = N_0$, $L(0) = L_0$, $C(0) = C_0$, $M(0) = M_0$, $I(0) = I_0$.

Ці рівняння мають загальні початкові додатні умови.

Нижче обговорюються функціональні форми, які вибираються для кожного члена

взаємодії клітин. Ми вирішили завжди позначати параметри малими літерами, що вважаються додатними.

1. dT/dt : **пухлинні клітини.**

У рівнянні (1) передбачається, що ріст пухлини є логістичним на основі даних, зібраних з мишей з імунодефіцитом, і представлений $aT(1-bT)$.

Ці члени представляють негативну взаємодію між двома популяціями. Вони можуть конкурувати за простір і поживні речовини, а також регулятивні дії та прямої взаємодії клітинної популяції. Взаємодія між пухлиною і NK-клітинами представлена членом масової дії $-cNT$. Лізис пухлини $CD8^+$ Т-клітинами, з іншого боку, має вигляд:

$$d \frac{(L/T)^I}{s + (L/T)^I} T.$$

Математична форма цього члену, що відображає обмежену природу здатності популяції ефекторних клітин лізувати пухлинні клітини, була визначена шляхом підгонки до опублікованих даних лабораторних мишей. $D(T, L) = d \frac{(L/T)^I}{s + (L/T)^I}$ представляє знищення пухлинних клітин Т-клітинами членом DT .

Модель містить член знищення при лікуванні у кожному з рівнянь клітинної популяції. Для простоти ми припускаємо кінетику дії маси, в якій кількість популяції клітини, що вбита лікуванням, пропорційну взаємодії клітин з ліками. Ми визнаємо, що деякі хіміотерапевтичні препарати, такі як доксорубіцин, ефективні лише упродовж певних фаз клітинного циклу. Фармакокінетика також показує, що ефективність хіміотерапії обмежена.

Як альтернативний член знищення клітини міг би розумно представляти нелінійний характер обмеженості відповіді. Однак для простоти ми вирішили включити у цій версії моделі лінійну форму дії мас. Тому ми представляємо вплив хіміотерапії на кожну клітинну популяцію як $K_\phi \phi$ для $\phi = T, N, L, C$. У випадку популяції пухлинних клітин член знищення хіміотерапією $-K_T MT$.

2. dL/dT : **пухлиноспецифічні ефекторні клітини, Т-клітини.** Клітинний ріст для $CD8^+$ Т-клітин складається лише з природних показників смертності. Швидкість загибелі Т-клітин вважається пропорційною популяції Т-клітин і представлена у рівнянні (3) як mL .

Інактивація цитолітичного потенціалу відбувається, коли ефекторна клітина (NK або $CD8^+$ Т-клітина) кілька разів взаємодіяла з пухлинними клітинами і перестає бути ефективною. Член інактивації позначено як $-qLT$. Параметр q являє собою середню швидкість інактивації.

$-uL^2$ описує саморегуляцію та пригнічення активності $CD8^+$ Т-клітин, що відбувається, коли у системі є дуже високі рівні активованих $CD8^+$ Т-клітин без чутливості до цитокінів. На відміну від моделі [p18] ми припускаємо, що саморегуляція відбувається через активність Т-лімфоцитів і не залучає NK клітини. Цей член входить у дію, коли в організмі висока кількість $CD8^+$ Т-клітин. Експериментальні дані підтверджують, що ці клітини можуть швидко інактивуватися навіть за наявності пухлини. Наявність пухлинних клітин також стимулює імунну систему виробляти більше $CD8^+$ Т-клітин. Припускаємо розпізнавання наявності пухлини пропорційно середньому числу зустрічей між циркулюючими лімфоцитами і пухлиною: $r_2 CT$. Активація $CD8^+$ включається за допомогою імунотерапії інтерлейкіном-2 (IL-2). Це фактично природний

цитокін в організмі людини та його ефект на ефективність імунної системи описується математично за допомогою члена взаємодії Міхаеліса-Ментена у рівнянні для L . Наявність IL-2 стимулює виробництво $CD8^+$ Т-клітин, що представляється членом $\frac{p_I I}{g_I + I} L$. Він має ту саму форму, що й член активації, [p14].

Член хіміотерапії представлено як $-K_L M L$. Член лікування лімфоцитами інфільтрації пухлини (TIL), $V_L(t)$, для популяції $CD8^+$ Т-клітин, являє собою імунотерапію, в якій рівень імунних клітин підвищується шляхом прямого додавання антиген-специфічного цитолітику імунних клітин.

3. dN/dt : неспецифічні ефекторні клітини, НК-клітини.

Припускається постійна швидкість виробництва НК-клітин і пропорційний популяції рівень смертності, представлений як $\alpha_1 - fN$ у рівнянні (2). Припущення про постійну швидкість виробництва НК відрізняється від моделі [p18], в якій швидкість виробництва НК пов'язана з популяціями циркулюючих лімфоцитів. У цьому випадку припущення про постійну продукцію НК також узгоджується з моделлю [p15], в якій також запропоновано цей параметр. Залучення НК-клітин представлено модифікованим законом взаємодії Міхаеліса-Ментена, за якого присутність пухлинних клітин стимулює вироблення і залучення додаткових неспецифічних ефекторних клітин. Звідси член рекрутування НК-клітин є

$$g \frac{T^\eta}{h + T^\eta} N.$$

Як і для пухлиноспецифічних лімфоцитів, для неспецифічних лімфоцитів додано член, що представляє їхню інактивацію після деяких зустрічей з пухлинними клітинами: $-K_N M N$. $-pNT$ представляє знищення неспецифічної клітинної популяції хіміотерапією.

4. dC/dt : циркулюючі лімфоцити.

Припускається, (4), що циркулюючі лімфоцити генеруються постійно і що кожна клітина має природний термін життя. Зростання і смерть представлені $\alpha_2 - \beta C$. Хіміотерапія вбиває популяцію циркулюючих клітин лімфоцитів, представлену як $-K_C M C$.

5. dI/dt : IL-2.

Відомо, що цитокін інтерлейкін-2 (IL-2) стимулює рекрутинг $CD8^+$ Т-клітин і проліферацію. IL-2 виробляється організмом природним шляхом, а також вводиться у терапевтичних цілях для підвищення функції імунної системи. IL-2 має використовуватися у клінічних дослідженнях окремо, у поєднанні з хіміотерапією та з ін'єкціями TIL (лімфоцити, що інфільтрують пухлину), у яких у систему вводиться велика кількість високоактивованих $CD8^+$ Т-клітин. У рівняння (5) додано члени, що представляють ендогенне виробництво інтерлейкіну-2 (IL-2) на додаток до можливості екзогенного застосування IL-2 як лікування. Перший член у рівнянні (5), $\frac{p_T T}{g_p T + T} L$ --

розуміння того, що взаємодія пухлиноспецифічних Т-клітин з пухлинними клітинами стимулює вироблення IL-2. Форма члену Міхаеліса-Ментена пояснюється самообмежувальним характером цього виробництва. Це представляє wLI . Передбачається, що концентрація IL-2 знижується експоненційно, що представлено як $-p_I I$. Нарешті, пряме екзогенне застосування IL-2 представлено залежним від часу

членом $V_I(t)$.

6. dM/dt : Хіміотерапевтичний препарат.

Припускається, що після застосування препарату хіміотерапії буде виведено з організму тіло з часом зі швидкістю, пропорційною його концентрації. Це представлено експоненційним членом розпаду, $-\gamma M$. Пряме застосування хіміотерапії включається через $V_M(t)$.

8.2.2 Вибір параметрів

Загальна модель з неспецифічними параметрами є корисною для вивчення якісної динаміки росту раку, а також може бути повчальною для виконання моделювання з метою спостереження за прикладами результатів.

Сила хіміотерапії вважалася рівною одному логарифму вбивства, і були обрані значення параметрів убивства K_T , K_N , K_M , K_C відповідно. Деякі хіміотерапевтичні препарати, такі як доксорубіцин, вбивають клітини, порушуючи процес поділу та росту. Клітини, що швидко діляться, як у волоссі, слизовій шлунку та кістковому мозку, де виробляються імунні клітини, переважно пошкоджуються хіміотерапією. Високі дози препарату також можуть пошкодити інші тканини в організмі. В ідеалі хіміотерапія повинна ефективніше за імунні клітини вбивати пухлинні клітини. Отже, K_T , K_N , K_M , K_C вважаються меншими за K_T , але величинами того самого порядку.

Швидкість розпаду препарату, γ , розраховували на основі періоду піврозпаду препарату та співвідношення $\gamma = \ln 2/t_{1/2}$. Період піввиведення препарату, $t_{1/2}$, оцінюється приблизно у 18 годин, що ґрунтується на експериментальних даних для хіміотерапевтичного препарату доксорубіцин.

Оцінено параметри p_I , g_I , p_T , g_T , w . Вибір перших чотирьох параметрів засновано на аналогічних значеннях параметрів, представлених у [pl4], але значення були дещо змінені, щоб відповідати поточній математичній моделі, яка відрізняється від моделі [pl4]. Період піврозпаду ефекторної клітини, $\ln 2/\mu_I$, є біологічна величина, що взято з моделі [pl4]. Оцінюються тільки значення r_2 і u . Ми помітили, що оскільки L , пухлиноспецифічна популяція $CD8^+$ Т-клітин, продукує I , популяцію IL-2 I продукує L , тобто, існує цикл зворотного зв'язку, що призводить до ефекту резонансу, отже, якщо його не контролювати, I зростає неконтрольовано. Тому нам довелося вибрати значення параметрів, які б давали відповіді у I та L за наявності пухлини, яка могла б утримувати популяцію IL-2 обмеженою.

w являє собою фракційну продукцію IL-2 з Т-клітин і його обрано так, щоб мати досить малу вартість. Це мале числове значення w допомагає уникнути ініціювання резонансного ефекту, а кількість IL-2 залишається на прийнятному рівні.

8.2.3 Приклади числової рівноваги

Щоб зрозуміти динаміку системи, необхідно ідентифікувати точки рівноваги та визначити їх характеристики стійкості.

Табл. 8.1. Розрахункові значення параметрів

Парам.	Опис	Значення	Розмірність
a	швидкість росту пухлини	$2 \cdot 10^{-3}$	1/доба
b	$1/b$ є місткість пухлиноносія	$1,02 \cdot 10^{-9}$	1/клітина
c_1	Фракційне знищення пухлинних клітин		
	ефекторними клітинами	$3,23 \cdot 10^{-7}$	1/(кл.доба)
d	рівень насичення фракційної пухлинної		
	клітини, що гинуть $CD8^+$ Т-клітинами	5	1/доба
f	смертність клітини-ефектора	0,0412	1/доба
g	максимальна ефекторна норма набору		
	по пухлинні клітини	0,025	1/доба
g_I	крива крутизни $CD8^+$ Т-клітин		
	набору по ІЛ-2	$2 \cdot 10^7$	IU/liter
g_T	крива крутизни ІЛ-2 набору від		
	пухлино-ефекторна клітина взаємодії	$1 \cdot 10^{57}$	клітин
h	коефіцієнт крутизни ефекторної клітини	600	клітин ^η
K_C, K_L	фракційні лімфоцити		
	що вбиті хіміотерапією	0,600	літр/(мг доба)
K_N	фракційні пухлинні клітини,		
	що вбиті хіміотерапією	0,800	літр/(мг доба)
l	показник дробу	2/3	
m	рівень смертності $CD8^+$ Т-клітин	0,02	1/доба
p	інактивація ефекторної швидкості		
	клітини пухлинними клітинами	$1,00 \cdot 10^{-7}$	1/(кл. доба)
p_I	максимум $CD8^+$ Т-клітин норма		
	набору по ІЛ-2	0,125	1/доба

Табл. 8.2. Розрахункові значення параметрів

Пар.	Опис	Значення	Розмірність
p_T	максимальна ІЛ-2 норма набору		
	від пухлино-ефекторна клітина взаємодії	0,60	$\frac{IU}{Л.кл.доба}$
q	швидкість інактивації $CD8^+$ Т-клітин		
	пухлинними клітинами	$3,42 \cdot 10^{-10}$	1/(кл. доба)
r_2	швидкість, за якою $CD8^+$ Т-клітини		
	стимулюються до вироблення в результаті		
	взаємодії пухлинних клітин лімфоцитами	$3 \cdot 10^{-11}$	1/(кл. доба)
s	коефіцієнт крутизни пухлина-		
	($CD8^+$ Т-клітина) лізису D	0,30	

u	масштабування саморегуляції Т-клітин	3,0	1/(кл. доба)
w	часткове виробництво ІЛ-2 Т-клітинами	$2 \cdot 10^{-4}$	1/(кл. доба)
α_1	постійне джерело ефektorні клітини	$1,3 \cdot 10^4$	1/(кл. доба)
α_2	постійне джерело циркулюючих лімфоцитів	$5 \cdot 10^8$	1/(кл. доба)
β	смертність кровообігу лімфоцитів	0,012	1/доба
γ	швидкість розпаду препарату хіміотерапії	0,90	1/доба
η	показник дробу рекрутування НК-клітин	1,0	
μ_I	швидкість розпаду препарату ІЛ-2	10,0	1/доба

Однак через аналітичне зменшення проблеми розв'язування шести рівнянь із шістьма змінними до задачі розв'язування одного рівняння з однією змінною (змінна T , що представляє популяцію пухлинних клітин), ми можемо отримати уявлення про певні динамічні характеристики, такі як можлива кількість точок рівноваги у системі. Весь аналіз не залежить від конкретного вибору значень параметрів.

8.2.4 Оптимальне управління

Отримаємо необхідні умови для оптимального управління мінімізації розміру пухлини з часом, а також мінімізації кількості введених ліків. Ми мінімізуємо дози ліків, оскільки припускаємо, що токсичні побічні ефекти викликають занепокоєння, і чим менша доза, тим краще. Ми мінімізуємо функціонал у формі, яка відображає потрібний компроміс для мінімізації як розміру пухлини, так і дози ліків:

$$J(V_L, V_I, V_M) = \int_0^{t_f} \left[T(t) + \frac{c_L}{2} V_L^2(t) + \frac{c_I}{2} V_I^2(t) + \frac{c_M}{2} V_M^2(t) \right] dt. \quad (197)$$

Для зручності позначимо $V(t) = [V_L(t), V_I(t), V_M(t)]$.

Необхідні умови оптимальності

Гамільтоніан:

$$\begin{aligned} H = & T + \frac{1}{2} \left[\varepsilon_L V_L^2(t) + \varepsilon_I V_I^2(t) + \varepsilon_M V_M^2(t) \right] \\ & + \lambda_1 \left[aT(1-bT) - c_1 NT - DT - K_T MT \right] + \lambda_2 \left[\alpha_1 - fN + \frac{gT^\eta}{h+T^\eta} N - pNT - k_N MN \right] \\ & + \lambda_3 \left[-mL - qLT - uL^2 + r_2 CT + \frac{p_I I}{g_I + I} L - K_L ML + V_L(t) \right] \\ & + \lambda_4 \left[\frac{p_T T}{g_T + T} L + wLI - \mu_I I + V_I(t) \right] + \lambda_5 (\alpha_2 - \beta C - K_C MC) + \lambda_6 (-\gamma M + V_M(t)) \end{aligned}$$

із $W_i(t) \geq 0$ -- штрафні функції:

Лагранжіан

$$\mathcal{L} = H + W_1(t)V_L(t) - W_2(1-V_L(t)) - W_3(t)V_M(t) - W_4(1-V_M(t)) - W_5(t)V_I(t) - W_6(1-V_I(t)). \quad (198)$$

$W_i(t) \geq 0$ — штрафні множники. Теорема 5.1. Враховуючи оптимальну трійку керування,

$$V(t) = (V_L^*(t), V_I^*(t), V_M^*(t))$$

розв'язків відповідної системи стану, існують приєднані змінні λ_i , $i = \overline{1,6}$, що задовольняють:

$$\begin{aligned} \frac{d\lambda_1}{dt} &= -\frac{\partial \mathcal{L}}{\partial T} = -1 - \lambda_1 \left[a - 2abT - c_1T - D + \frac{sdl(L/T)^l}{[s + (L/T)^l]^2} - K_T M \right] \\ &\quad + \lambda_2 \left[g \frac{h\eta T^{\eta-1}}{(h+T^\eta)^2} N - pN \right] - \lambda_3(-qL + r_2C) + \lambda_4 \frac{p_T q_T}{(g_T + T)^2} L; \\ \frac{d\lambda_2}{dt} &= -\frac{\partial \mathcal{L}}{\partial N} = \lambda_1 c_1 T - \lambda_2 \left(-f + g \frac{T^\eta}{h+T^\eta} - pT - K_N M \right); \\ \frac{d\lambda_3}{dt} &= -\frac{\partial \mathcal{L}}{\partial L} = \lambda_1 \frac{sdl(L/T)^{l-1}}{[s + (L/T)^l]^2} - \lambda_3 \left(-m - qT - 2uL + \frac{p_I I}{g_I + I} - K_L M \right) - \lambda_4 \left(\frac{p_T T}{g_T + T} + wI \right); \\ \frac{d\lambda_4}{dt} &= -\frac{\partial \mathcal{L}}{\partial I} = -\lambda_3 \frac{p_I g_I}{(g_I + I)^2} L - \lambda_5(wL - \mu_I); \\ \frac{d\lambda_5}{dt} &= -\frac{\partial \mathcal{L}}{\partial C} = -\lambda_3 r_2 T - \lambda_5(-\beta - K_C M); \\ \frac{d\lambda_6}{dt} &= -\frac{\partial \mathcal{L}}{\partial M} = \lambda_1 K_T T + \lambda_2 K_N N + \lambda_3 K_L L + \lambda_5 K_C C + \gamma \lambda_6, \end{aligned}$$

де $\lambda_i(t_f) = 0$, $i = \overline{1,6}$.

$V_L^*(t), V_I^*(t), V_M^*(t)$ можна подати як

$$\begin{aligned} V_L^* &= \min \left[1, \left(\frac{-\lambda_3}{\varepsilon_L} \right)^+ \right]; \\ V_M^* &= \min \left[1, \left(\frac{-\lambda_6}{\varepsilon_M} \right)^+ \right]; \\ V_I^* &= \min \left[1, \left(\frac{-\lambda_4}{\varepsilon_I} \right)^+ \right]; \end{aligned}$$

де $r^+ = 1(r)r$.

Наприклад, управління V_M^* показує, що кандидат на оптимальний курс хіміотерапії набагато складніший за поточні стандартні схеми лікування, які проводяться у клініці. Ці стандартні режими, як правило, дотримуються шаблону із кількох днів увімкнення, потім кількох вихідних, залежно від того, яку кількість ліків може перенести пацієнт, і чи перервуть лікування вихідні.

Однак оптимальним режимом, згідно з нашим аналізом, був би не тільки залежність від часу, але його буде додатково модифіковано як функція сили та активності усіх чотирьох популяцій клітин у моделі.

8.2.5 Чисові результати застосування квадратичного керування

Використовуються необхідні умови оптимальності першого порядку як цільовий функціонал. Вибір початкових умов зазвичай дає лише локально оптимальні рішення. Моделюються два типи сценаріїв: перший, у якому ймовірна початкова кількість пухлинних клітин буде клінічно непомітною ($T = 100$, називається «малим») і другим, в якому кількість пухлинних клітин знаходиться на вершині виявлення ($T = 10^7$, позначається як «виявлений»). Протестовано два набори початкових умов, «малі» і «виявлені» є такими. Малі початкові умови є

$$T = 100, N = 3 \cdot 10^4, L = 30, I = 500, C = 5 \cdot 10^6, M = 0.$$

$$T = 10^7, N = 5 \cdot 10^5, L = I = 2 \cdot 10^3, C = 4,17 \cdot 10^{10}, M = 0.$$

Вагові коефіцієнти:

$$1. \varepsilon_I = 10^7, \varepsilon_L = 0, \varepsilon_M = 10^7.$$

$$2. \varepsilon_I = 10^5, \varepsilon_L = 10^5, \varepsilon_M = 10^7.$$

Без управління цей «маленький» набір початкових умов призводить до появи пухлини, що втягнута у півстіжку нульову рівновагу.

8.3 Моделювання виникнення і росту пухлин - карцинома

8.3.1 Вступ

Розглядаються питання виникнення однієї з розповсюджених онкологічних хвороб, карциноми. Існує два шляхи виникнення генетичної аномалії: 1) стимулюючий ген стає гіперактивним -- онкоген; 2) інгібуючий ген стає неактивним; він називається антінкоген, наприклад, ген p53, що контролює початок клітинного циклу. Коли один або кілька генів мутують і стають патологічними, клітина та її нащадки розмножуються набагато швидше за нормальні клітини. Новоутворення, або пухлина, являє собою збільшення патологічних клітин. Доки ця кількість обмежена порожниною, пухлина вважається доброякісною. Якщо пухлина виходить із порожнини яєрез базальну мембрану і потім розповсюджується на позаклітинному матриксі або стромі, вона стає злоякісною. Далі настає метастазування через кровотік або лімфатичні судини. Усім раковим клітинам потрібна більша кількість кисню за наявну у здоровій тканині. Тому вони продукують фактор росту ендотелію судин (VEGF) для росту і задіявання кровоносних судин. Такий процес називається ангіогенезом.

8.3.2 Загальні особливості математичних моделей ракових пухлин

Структура моделі

Тканина, де проліферують клітини, називається позаклітинним матриксом (ECM). Він має складну структуру, тому більшість математичних моделей спрощують структуру ECM.

Пористе середовище. Вважається, що пухлина складається з трьох типів клітин: проліферуючих клітин із густиною $p(x,t)$, клітин у спокої з густиною $q(x,t)$ та мертвих клітин із густиною $n(x,t)$.

Припускається, що клітини у спокої стають проліферуючими зі швидкістю $K_p(c)$, яка залежить від концентрації поживних речовин c і вони ж відмирають зі швидкістю $K_d(c)$. Аналогічно, проліферуючі клітини переходять у стан спокою зі швидкістю $K_q(c)$, а коефіцієнт їхньої загибелі дорівнює $K_a(c)$. Густина проліферуючих клітин збільшується зі швидкістю $K_b(c)$. Вважається, що мертві клітини виводяться із пухлини (за допомогою макрофагів) зі сталою швидкістю K_r .

Нехай усі клітини фізично ідентичні за об'ємом і масою, а їхня густина є стала по всій пухлині. Тоді

$$p + q + n = \text{const} = \theta. \quad (199)$$

Внаслідок проліферації та виведення клітин здійснюється їхній неперервний рух всередині пухлини. Цей рух опишемо полем швидкостей v . Тоді закон збереження маси для густини проліферуючих клітин p , клітин у спокої q та мертвих клітин n у межах області пухлини $\Omega(t)$ набуває вигляду

$$\frac{\partial p}{\partial t} + \text{div}(pv) = [K_b(c) - K_q(c) - K_a(c)]p + K_p(c)q; \quad (200)$$

$$\frac{\partial q}{\partial t} + \text{div}(qv) = K_q(c)p - [K_p(c) - K_d(c)]q; \quad (201)$$

$$\frac{\partial n}{\partial t} + \text{div}(nv) = K_a(c)p + K_d(c) - K_r n. \quad (202)$$

Пухлинна тканина розглядатиметься як пористе середовище, а рухомі клітини -- як потік рідини. У пористому середовищі швидкість потоку рідини v зв'язана із тиском рідини σ згідно закону Дарсі

$$v = -\nabla \sigma. \quad (203)$$

Якщо додати рівняння (10)--(12) і скористатися (9), матимемо

$$\theta \text{div} v = K_b(c)p - K_r n, \quad (204)$$

і можна замінити (12) на (14).

Вважається, що концентрація поживних речовин відповідає рівнянню дифузії

$$\varepsilon_0 \frac{\partial c}{\partial t} = \Delta c - \lambda(p + q) \text{ в } \Omega(t). \quad (205)$$

Виключимо n з (14) і покладемо для спрощення $\theta = 1$.

$$\frac{\partial p}{\partial t} - \nabla \sigma \nabla p = f(c, p, q) \text{ в } \Omega(t), t > 0, \quad (206)$$

$$\frac{\partial q}{\partial t} - \nabla \sigma \nabla q = g(c, p, q) \text{ в } \Omega(t), t > 0, \quad (207)$$

$$\Delta \sigma = -h(c, p, q) \text{ в } \Omega(t), t > 0, \quad (208)$$

де

$$f(c, p, q) = [K_b(c) - K_q(c) - K_a(c)]p + K_p(c)q - h(c, p, q)p,$$

$$g(c, p, q) = K_q(c)p - [K_p(c) - K_d(c)]q - h(c, p, q)q,$$

$$h(c, p, q) = -K_r + [K_b(c) + K_r]p + K_e q.$$

Позначимо $\Gamma(t)$ -- межа області $\Omega(t)$. Запроваджуються межові умови

$$c = \bar{c} \text{ в } \Gamma(t), t > 0, \quad (209)$$

$$\sigma = \gamma \kappa \text{ в } \Gamma(t), t > 0, \quad (210)$$

де $\bar{c}$ -- стала, γ -- коефіцієнт поверхневого натягу, що є міжклітинна адгезія; κ --

середня кривизна ($\kappa = 1/R$ коли $\Omega(t)$ -- шар радіусу R). Припускається, що нормальна швидкість V_n межі $\Gamma(t)$ дорівнює нормальній швидкості $v \cdot n$ клітин у зовнішньому нормальному напрямку n , тобто

$$\frac{\partial \sigma}{\partial n} = V_n \text{ на } \Gamma(t), t > 0, \quad (211)$$

Рідинна тканина. У разі злоякісної пухлини, що виникає у молочній залозі або у мозку, слушно вважати тканину рідинною і використовувати рівняння Стокса замість закона Дарсі. У цьому випадку співвідношення між швидкістю v і тиском σ визначається як

$$-\nu \nabla^2 v + \nabla \sigma = f, \quad (212)$$

$$\operatorname{div} v = g, \quad (213)$$

де $f = -\nu/3 \nabla g$ і межева умова (20) замінюється на

$$T(v, \sigma)n = -\gamma \kappa n, \quad (214)$$

де

$$T(v, \sigma) = \nu(\nabla v + \nabla v^T) - (\sigma + \frac{2}{3} \nu \operatorname{div} v)I.$$

Тут ν -- коефіцієнт в'язкості; I -- одинична матриця; g -- швидкість проліферації, що визначається правою частиною (14).

Густина позаклітинного матриксу (ЕСМ). Структура ЕСМ являє собою складну мережу білкових волокон, колагенів. Загальним спрощеним підходом є запровадження концентрації цих волокон ρ . Зазвичай волокна продукуються клітинами фібробластів. Ракові клітини (із густиною n) можуть рухатися всередині ЕСМ і руйнувати його волокна. Вони руйнують матрикс за допомогою ферменту, який має назву матриксної металопротеїнази (ММР). Постулюється, що ρ задовольняє рівнянню

$$\frac{\partial \rho}{\partial t} = -a_1 P n + a_2 f (1 - \rho/\rho_*), \quad (215)$$

де P -- концентрація ММР; ρ_* -- максимальна густина ЕСМ і a_i -- додатні сталі.

Рівняння для P і f залежать від конкретного раку і його мікрооточення.

Моделювання ангиогенезу

Під час зростання пухлини клітини у ядрі перестають отримувати досить кисню із судинної системи. На цій стадії ракові пухлини починають виділяти ангиогенний фактор пухлини (ТАФ), такий як VEGF (васкулярний ендотеліальний фактор росту), який стимулює ендотеліальні клітини виходити із існуючих капілярів, проліферувати і формувати нові капіляри (процес зветься ангиогенезом), що рухаються до пухлини.

Покладемо n -- густина клітин пухлини; e -- густина ендотеліальних клітин; a -- концентрація ТАФ; c -- концентрація протеолітичних ферментів (протеази); c_a -- активні форми c ; f -- концентрація фібронектину (це волокна, що зв'язують клітини із колагеном, що надає можливість клітинам рухатися у тканині). Тоді постулюється

$$c_a = \frac{c}{1 + \alpha_0 f}, \quad (216)$$

$$\frac{\partial n}{\partial t} = D_a \Delta a + \alpha_1 n - \frac{\alpha_2 a}{1 + \alpha_3 a} e, \quad (217)$$

$$\frac{\partial c}{\partial t} = D_c \Delta c + \frac{\alpha_2 a}{1 + \alpha_3 a} e - \alpha_4 c, \quad (218)$$

$$\frac{\partial f}{\partial t} = D_f \Delta f + \alpha_5 f \left(1 - \frac{f}{f_0} \right) - \frac{\alpha_6 f}{1 + \alpha_7 f} c_a, \quad (219)$$

$$\frac{\partial e}{\partial t} = D_e \nabla \left[\nabla \ln \frac{e}{\tau(c_a, f)} \right], \quad (220)$$

де

$$\tau(c_a, f) = \left(\frac{c_a + \beta_1}{c_a + \beta_2} \right)^{\gamma_1} \left(\frac{f + \delta_1}{f + \delta_2} \right)^{\gamma_2} \quad (221)$$

називається хемотаксичною функцією; $\alpha_i, \beta_i, \delta_i$ -- сталі.

Коефіцієнти D_c, D_f дуже малі і фактично не проявляються. Система рівнянь (26)--(31) має бути доповнена межовими і початковими умовами.

Ракові клітини отримують більше кисню поблизу межі й, отже, розмножуються швидше за клітини у ядрі. Більша частина проліферації здійснюється у зовнішніх оболонках пухлини, у той час як внутрішнє ядро в основному складається із мертвих клітин.

Клітковий цикл

Більшість математичних моделей ґрунтується на диференційних рівняннях, в яких шукані функції -- це густини клітин, концентрація білків чи інших молекул. Оскільки злоякісні пухлини виникають із-за мутації генів, моделі конкретного онкологічного захворювання мають враховувати явище, що має місце на мікроскопічному рівні. Крім того, "рішення" клітини проліферувати чи залишатися у спокої стається у конкретний момент часу клітинного циклу, як і рішення клітини здійснити самовбивство (апоптоз) у випадку виявлення пошкодження, яке неможливо виправити. Отже, поряд із часом t необхідно враховувати час клітинного циклу s , який зазвичай складає кілька діб. Клітковий цикл поділяється на чотири фази: S (синтез), M (мітоз) і фази розриву G_1 і G_2 . Під час S -фази ДНК реплікується, тобто кожна хромосома подвоюється. Під час другої фази (мітозу) ядерна мембрана руйнується і сестринські хроматиди відокремлюються, утворюючи нові ядерні мембрани, і клітина ділиться на дві. Фази S і M розділяються двома фазами розриву -- G_1 і G_2 . Клітинним циклом управляють дві контрольні точки: R_1 , що міститься на кінці G_1 і R_2 -- у G_2 .

У контрольній точці R_1 клітина вибирає один із трьох варіантів: 1) здійснити самогубство (апоптоз), якщо виявилось, що вона була пошкоджена без можливості відновлення під час фази росту G_1 ; 2) увійти у фазу спокою G_0 і залишатися там деякий час, якщо мікросередовище гіпотоксичне або перенаселене іншими клітинами; або 3) перейти до S -фази. У точці R_2 клітини вирішують або перейти в апоптоз, якщо непоправимий збиток стався під час реплікації ДНК, або перейти у M -фазу.

Багатомасштабна модель описується коли єдиними клітинами є ракові клітини із можливими мутаціями у наборі генів $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_l)$. Рівняння (10)--(12) мають бути

змінені відносно рівнянь для ракових клітин p_i у різних фазах i .

Запроваджуються наступні позначення:

$p_1(x, t, s_1)$ -- густина клітин у фазі G_1 ; $0 \leq s_1 \leq A_1$;

$p_2(x, t, s_2)$ -- густина клітин у фазі S ; $0 \leq s_2 \leq A_2$;

$p_3(x, t, s_3)$ -- густина клітин у фазах G_2 і M ; $0 \leq s_3 \leq A_3$;

$p_0(x, t, s_0)$ -- густина клітин у фазі G_0 ; $0 \leq s_0 \leq A_0$;

$p_4(x, t)$ -- густина мертвих клітин.

Нехай $w(x, t)$ -- концентрація кисню, а $Q(x, t)$ -- сумарна густина живих клітин у фазах G_1, S, G_2, M і G_0 . Тоді закон збереження маси має вигляд

$$\frac{\partial p_i}{\partial t} + \frac{\partial p_i}{\partial s_i} + \text{div}(p_i v) = \lambda_i(w) p_i, \quad 0 < s_i < A_i \quad (i = 1, 2, 3); \quad (222)$$

$$\frac{\partial p_0}{\partial t} + \frac{\partial p_0}{\partial s_0} + \text{div}(p_0 v) = \lambda_0 p_0, \quad 0 < s_0 < A_0; \quad (223)$$

$$\frac{\partial p_4}{\partial t} + \text{div}(p_4 v) = \mu_2 p_2(x, t, A_2) + \mu_1 p_1(x, t, A_1) - \lambda_4 p_4; \quad (224)$$

де $\lambda_i(w)$ -- швидкості зростання і $\mu_i = \mu_i(\gamma)$. Отже, густина клітин не змінюється під час реплікації:

$$p_1(x, t, 0) = p_3(x, t, A_3). \quad (225)$$

$$p_2(x, t, 0) = [1 - \mu_1 - K(w(x, t), Q(x, t); \gamma)] p_1(x, t, A_1) + p_0(x, t, A_0). \quad (226)$$

Перший член у правій частині рівняння ((226)) є рішення клітини у контрольній точці R_1 переходити до апоптозу (μ_1), до форми спокою або до S -фази; рішення залежить від середовища (w, Q) і від стану генів в γ . Останній член є клітини, що перейшли від фази G_0 до S .

Рішення клітини про перехід в апоптоз у точці R_2 описується рівнянням

$$p_3(x, t, 0) = (1 - \mu_2) p_2(x, t, A_2), \quad (227)$$

або залишатися у фазі спокою

$$p_0(x, t, 0) = K[w(x, t), Q(x, t), \gamma] p_1(x, t, A_1). \quad (228)$$

Система (32)--(38) ще не повна, оскільки не указана швидкість v . Для цього запроваджуються величини

$$Q_i(x, t) = \int_0^{A_i} p_i(x, t, s_i) ds_i \quad (i = \overline{0, 3},$$

$$Q_4(x, t) = p_4(x, t); \quad Q(x, t) = \sum_{i=1}^3 Q_i(x, t).$$

Із використанням (35)--(38) отримаємо

$$\sum_{i=0}^4 \left[\frac{\partial Q_i}{\partial t} + \text{div}(Q_i v) \right] = \sum_{i=1}^3 \lambda_i(w) Q_i - \lambda_0 Q_0 - \lambda_4 Q_4. \quad (229)$$

Якщо тепер припустити, що

$$\sum_{i=0}^4 Q_i(x, t) \equiv \text{const} = \theta, \quad (230)$$

із (39) впливає

$$\theta \operatorname{div} v = H(Q, w), \quad (231)$$

$$H(Q, w) = \sum_{i=1}^3 \lambda_i(w) Q_i - \lambda_0 Q_0 - \lambda_4 Q_4. \quad (232)$$

Концентрація кисню задовольняє рівняння дифузії

$$\frac{\partial w}{\partial t} - D_w \Delta w + \bar{\lambda} Q w = b, \quad (233)$$

де $\bar{\lambda}$ -- додатна стала, $b = 0$ у період аваскулярної стадії пухлини.

Тепер можна переходити до розгляду повної моделі за припущення, що тканина підпорядковується або закону Дарсі, або рівнянню Стокса.

Припускається, що тканину можна описати як пористе середовище, що складається з трьох фаз: клітин, позаклітинної рідини і шорсткого позаклітинного матриксу.

Розглянемо пухлину, розташовану у скінченній області $\Omega(t)$ площини $\Omega = \{x, y \mid 0 \leq x \leq x_{\max}, 0 \leq y \leq t_{\max}\}$.

Пухлинна область $\Omega(t)$ зростає із часом t . Доповнення $\Omega_h(t) = \Omega \setminus \Omega_t(t)$ вважається заповненою тканиною хазяїна.

Нехай: $\varphi(x, t)$ -- об'ємна частка клітин; $v = v(x, t)$ -- поле швидкостей фази клітини; $\varphi_l(x, t)$, $v_l(x, t)$ -- об'ємна частка і швидкість позаклітинної рідинної фази; φ_{ECM} -- фіксована частка шорсткого позаклітинного матриксу. $x = (x, y)$ -- точка на площині. Маємо:

$$\varphi + \varphi_l + \varphi_{ECM} = 1 \rightarrow \varphi + \varphi_l + \varphi_* = \text{const}. \quad (234)$$

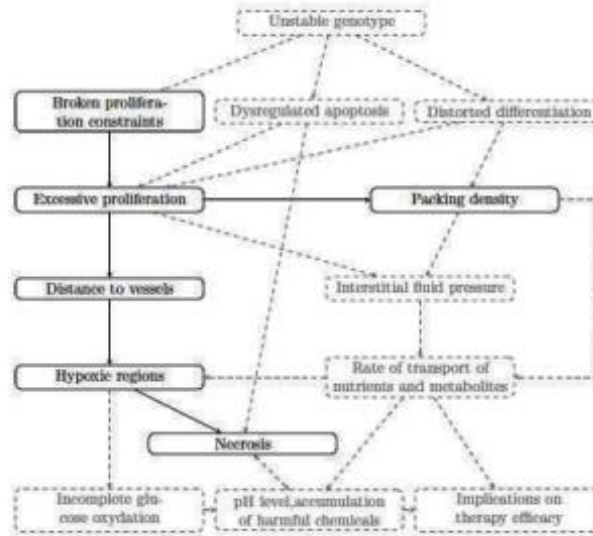


Рис.8.3 Область застосування моделі. Пунктирними лініями позначено деякі явища, що є важливі для розвитку пухлинного шнура і терапії пухлин, але не охоплюються моделлю

Баланс маси і сил.

$$\frac{\partial \varphi}{\partial t} + \operatorname{div}(\varphi v) = \frac{\Gamma}{\rho}, \quad (235)$$

$$\frac{\partial \varphi_l}{\partial t} + \operatorname{div}(\varphi_l v) = \frac{\Gamma_l}{\rho}, \quad (236)$$

де Γ, Γ_l -- швидкості росту клітин і рідинної фази відповідно. $\rho = \text{const}$.

8.3.3 Гліома

Гліома характеризується інвазивністю в оточуючі нормальні тканини мозку, що призводить до рецидиву пухлини після операції. Пацієнти з діагностованою гліобластою, яка є найагресивніша із пухлин мозку, мають термін виживаємості приблизно один рік із моменту діагностування. На міграцію клітин гліоми у головному мозку впливає кілька факторів. Міграція може стимулюватися позаклітинним матриксом (ЕСМ) у процесі гаптотаксичної міграції. Гаптотаксис вмикається через існуючі компоненти мозку.

У пацієнтів клітини гліоми зазвичай слідують вздовж білої речовини і базальної пластинки кровоносних судин головного мозку.

Розглянемо математичну модель для клітин гліоми *in vitro* (у пробірці) поза сферичного ядра, яка описує міграцію, розкидання і розгалуження клітин, що спостерігаються в експериментах. Модель містить матриксні металопротеїнази і концентрацію поживних речовин і враховує ефекти хемотаксису, гаптотаксису і позаклітинної адгезії.

Нехай n -- густина клітин гліоми; ρ -- концентрація позаклітинного матриксу; P == концентрація матриксної металопротеїнази; G -- концентрація поживних речовин.

Область міграції -- це оболонка $\Omega = \{R_0 < r < R_1\}$, що охоплює сферичну пухлину фіксованого радіусу R_0 . Справедливі такі рівняння:

$$\begin{aligned} \frac{\partial n}{\partial t} = D_n \Delta n - \nabla \left(\chi_n \frac{n \nabla G}{\sqrt{1 + \lambda_G |\nabla G|^2}} \right) \\ - \nabla \left(\chi_n^1 \frac{n \nabla \rho}{\sqrt{1 + \lambda_\rho |\nabla \rho|^2}} \right) - \nabla (nK(n)) + \lambda_{11} n + S; \end{aligned} \quad (237)$$

$$\frac{\partial \rho}{\partial t} = -\lambda_{21} P \rho + \lambda_{22} \rho \left(1 - \frac{\rho}{\rho_0} \right); \quad (238)$$

$$\frac{\partial P}{\partial t} = D_P \Delta P + \lambda_{31} n - \lambda_{32} P; \quad (239)$$

$$\frac{\partial G}{\partial t} = D_G \Delta G - \lambda_{41} n \frac{G}{G + k_G}. \quad (240)$$

Тут $n(x, t)K(n(x, t))$ -- сила зчеплення, що діє на $n(x, t)$;

$$K[n(x, t)] = \frac{2}{|B_\delta(x)|} \int_{B_\delta(x)} \lambda_a f(n(y, t)) S(y) dy,$$

$$S(y) = \frac{y - x}{|y - x|}, \quad f(n) = k_s \frac{n}{n_0 + n},$$

$k_s, n_0, \lambda_a, \eta$ -- параметри.

Функція розкидання S є унікальна для гліоми. Вона описує як ракові клітини викидаються випадковим чином з поверхні пухлини.

Досягнення в області вірусології та біології пухлин надали можливість створити

онколітичні віруси (ОВ), які вибірково реплікуються у пухлинних клітинах. Нащадки ОВ розповсюджуються по всій пухлині без пошкодження нормальних клітин.

Стандартне лікування пацієнтів з гліомою -- операція із подальшою променевою терапією. Але продовжуються експерименти лікування із застосуванням ОВ, в яких частина клітин гліоми впроваджується у мозок миші. При цьому уроджена імунна система знищує ракові клітини, що заражені вірусом, зменшуючи кількість вільного вірусу і ефективність терапії.

Математична модель використовує такі змінні: x -- концентрація ракових клітин; y -- концентрація інфікованих (заражених) ракових клітин; n -- концентрація мертвих клітин; v -- концентрація вільних вірусів, що не містяться у ракових клітинах; z -- концентрація імунних клітин, що атакують інфіковані пухлинні клітини.

Вважається, що пухлина радіально симетрична і через $u(x,t)$ позначено радіальну швидкість, $u(x,0) = 0$. Тоді при збереженні концентрації в області пухлини $r < R(t)$ виконуються такі рівняння:

$$\frac{\partial x}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 ux) = \lambda x - \beta xv, \quad (241)$$

$$\frac{\partial y}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 uy) = \beta xv - kyz - \delta y, \quad (242)$$

$$\frac{\partial n}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 un) = kyz - \omega z^2 - \mu n, \quad (243)$$

$$\frac{\partial z}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 uz) = syz - \omega z^2 - P(t), \quad (244)$$

$$\frac{\partial v}{\partial t} - D \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial v}{\partial r} \right) = b \delta y - k_0 vz - \gamma v. \quad (245)$$

Тут λ -- коефіцієнт проліферації пухлинних клітин; β -- коефіцієнт зараження; δ -- коефіцієнт лізису інфікованих клітин; μ -- коефіцієнт видалення мертвих клітин; D -- коефіцієнт дисперсії вірусів; b -- розмір розриву інфікованих клітин; γ -- коефіцієнт очищення вірусу; k -- коефіцієнт імунних убивств; k_0 -- коефіцієнт поглинання вірусів; s -- коефіцієнт стимуляції інфікованих клітин; ω -- коефіцієнт очищення імунних клітин; $P(t)$ -- концентрація циклофосаміду.

Друга складова у правій частині рівняння (51) -- це хемотаксис, третя -- гаптотаксис, 4-а -- самозлипання, 5-а -- проліферація, S -- розкид. Перша складова у правій частині рівняння (52) -- руйнування, друга -- відновлення.

Модель передбачає, що якщо розмір розриву $b > 100$, радіус ракової пухлини буде зменшуватися. Наведена модель не враховує побічні ефекти пригнічення імунітету.

Розділ 9 Лікування гліоми

9.1 Вступ

Онколітичні віруси -- генетично змінені, здатні до реплікації віруси, які інфікують ракові клітини та розмножуються в них, але не завдають шкоди нормальним клітинам. Коли заражена клітина гине, вивільняються і поширюються багато новоутворених вірусів і заражають сусідні пухлинні клітини. Проведено лікування онколітичними вірусами (ОВ) і продовжує активно тестуватися в клінічних випробуваннях на різні види раку із використанням різноманітних вірусів [R2], враховуючи ONYX-15, вірус простого герпесу (HSV) і простатоспецифічний аденовірус CN706 і CN707.

Ця терапія, хоча і заснована на досить багатообіцяючих припущеннях, стикається з однією головною перешкодою: вроджена імунна система розпізнає інфіковані клітини та знищує їх до того, як віруси в них отримають шанс розмножитися.

Мета полягає у тому, щоб побудувати модель, яка фіксує взаємодію між здоровими пухлинними клітинами, інфікованими пухлинними клітинами, вірусами та макрофагами та $TNF-\alpha$, який вони виробляють. Модель PDE базується на [R2]. Але тут ми сформулюємо модель ODE популяції. Це скорочення дозволить нам провести детальний динамічний системний аналіз моделі, а також аналіз її як задачі оптимального керування.

Ефективність обох методів лікування шляхом ін'єкції вірусу та інгібітору $TNF-\alpha$ аналізуватиметься у контексті радіуса пухлини, повертаючи просторовий аспект.

Підхід полягатиме в націлюванні на пухлину шляхом комбінування двох терапій: вірусної ін'єкції та інгібітору $TNF-\alpha$. Будемо аналізувати реакцію системи на різні дози, зокрема ефективність терапії, маючи на меті мінімізацію радіуса пухлини. Взаємодію із вірусом буде досягатися шляхом неперервного введення фіксованої дози. Це є мотивовано тим, що терапевтична ефективність системних хіміопрепаратів суттєво обмежена через наявність високоселективних мозкових гематоматозних бар'єрів (ГЕБ) (BBB). Щоб подолати це обмеження, дослідники розробили новий метод доставки, розширена конвенційна доставка (CED), яка спирається на внутрішньочерепну доставку вірусних векторів шляхом неперервної інфузії через катетери. CED нещодавно досяг фази I/II у клінічних випробуваннях, хоча досі ця процедура було показано лише помірну ефективність у пацієнтів з рецидивом гліоми.

Обидва терапевтичні засоби, вірус та інгібітор $TNF-\alpha$, мають негативні побічні ефекти. Побічні ефекти вірусу, такого як HSV, який використовується у вірусотерапії пухлин, є значними на моделях пухлин гризунів, і невідомо, чи вони відбуватимуться також у людей [R1]. У випадку інгібіторів $TNF-\alpha$ побічні ефекти включають опортуністичні інфекції, такі як туберкульоз, лістероз і пневмоцистоз. Бувають і побічні явища у тому числі системний вовчак, застійна серцева недостатність та загальні інфекції [R1]. Через ці побічні ефекти кількість дозування має бути оптимізована для максимізації ефекту та мінімізувати негативні побічні ефекти.

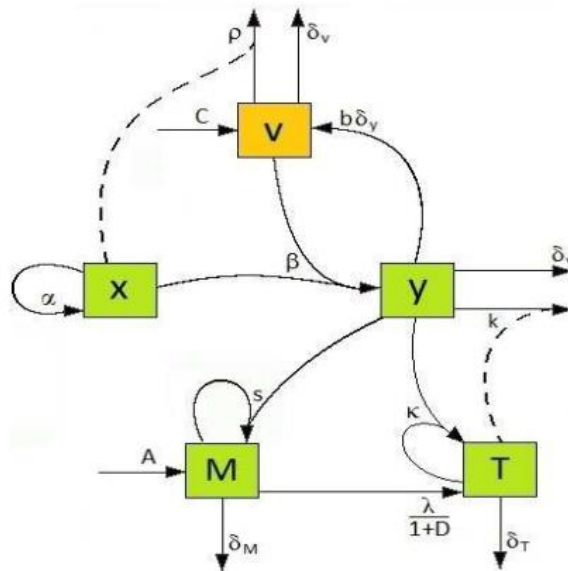


Рис. 9.1 Блок-схема моделі для віротерапії

Мультиформна гліобластома (GBM) є найруйнівніша пухлина мозку. Незважаючи на агресивну терапію, що включає резекцію пухлини, променеву та хіміотерапію, середній час виживання після постановки діагнозу становить 15 місяців. Онколітичний вірус (ОВ) є генетично змінений вірус, який може вибірково розмножуватися в пухлинних клітинах і вбивати їх, при цьому не завдаючи шкоди нормальним, здоровим клітинам, [R2].

Онколітична вірусна терапія була досліджена як підхід до боротьби з агресивними пухлинами, включаючи гліобластоми, [R8]. Одним із факторів, що обмежує терапію ОВ, є антигенність інфікованих пухлинних клітин; імунна система розпізнає ці клітини і знищує їх разом із вірусами всередині. Іншим фактором, що знижує ефективність терапії ОВ, є адгезивність позаклітинного матриксу (ECM) у мікрооточенні пухлини, що уповільнює вільний рух позаклітинних вірусів. Лікування гліоми хондроїтиназою ABC I (Chase ABC), щоб «послабити» каркас ECM і підвищити ефективність ОВ терапії, досліджували в [R3,R4].

Ще одним фактором, який обмежує ефективність лікування ОВ при гліомі, є CCN1 , протеїн позаклітинного матриксу, який, як відомо, бере участь у ангіогенезі, клітинній міграції та запаленні, [R5,R6], CCN1 посилюється під час інфікування гліоми онколітичним простим герпесом (HSV-1).

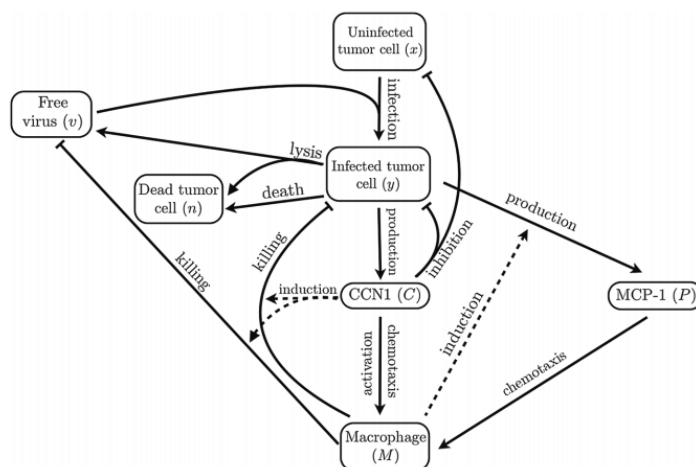


Рис. 9.2 Модель мережі взаємодії

Неінфіковані пухлинні клітини заражаються при проникненні вільного вірусу. Інфіковані клітини піддаються лізису, що призводить до вивільнення вільних вірусів. Інфіковані клітини виробляють білки CCN1 і MCP-1, які є хемоаттрактантами для макрофагів. CCN1 керує протівірусною відповіддю, а також пригнічує ріст пухлини. Крім того, CCN1 викликає прозапальну активацію макрофагів і індукує загибель інфікованих клітин і вірусу макрофагами. Макрофаги, активовані CCN1, сигналізують про інфіковані клітини гліоми, щоб збільшити виробництво MCP-1.

9.2 Віротерапія гліоми

9.2.1 Математична модель лікування гліоми віротерапією та інгібіторами TNF- α

Нехай $x(t)$ позначає густину ракових клітин (неінфікованих), $y(t)$ -- густину інфікованих ракових клітин, $v(t)$ -- густину вірусу, $M(t)$ -- густину макрофагів, а $T(t)$ -- концентрація TNF- α . Модель містить два елементи керування, u_1 і u_2 . $u_1(t)$ представляє кількість вірусу, який вводиться в пухлину, а $u_2(t)$ означає дозування інгібітору TNF- α . "in" і "out" протікає між відсіками та дією двох препаратів, що схематично показано на рис. 8.1.

Число спалахів — це кількість вірусів, які виходять із вмираючої ракової клітини. Візьмемо його у діапазоні $50 \leq b \leq 150$. Параметр b є розмір спалаху, визначений як $b \times \frac{\text{massofvirus}}{\text{massofcells}} = b \cdot 10^{-6} \text{ (g/cm}^3\text{)}$.

Модель динаміки описується такою системою ОДУ:

$$\frac{dx}{dt} = \alpha x - \beta xv - \delta_x x, \quad (1)$$

$$\frac{dy}{dt} = \beta xv - ky \frac{T}{K+T} - \delta_y y, \quad (2)$$

$$\frac{dM}{dt} = A + syM - \delta_M M, \quad (3)$$

$$\frac{dT}{dt} = \frac{\lambda}{1+u_2} M - ky \frac{T}{K+T} - \delta_T T, \quad (4)$$

$$\frac{dv}{dt} = b_1 ky \frac{T}{K+T} + b \delta_y y - \rho xv - \delta_v v + u_1. \quad (5)$$

Усі густини та концентрації вказані в одиницях g/cm^3 . У рівнянні (1) α представляє швидкість проліферації неінфікованих ракових клітин і δ_x — швидкість смертності; β є швидкість інфікування x вірусами v . У (2) $ky \frac{T}{K+T}$ означає некротичну загибель інфікованих клітин, спричинену TNF- α , тоді як δ_y — загибель через апоптоз. Коли клітина y гине шляхом апоптозу, вірусні частинки b вивільняються, а якщо вона гине шляхом некрозу, виникає дуже невелика кількість, b_1 , вірусів. Вони враховуються першими двома членами у правій частині рівняння (5). У рівнянні (3) складові A і $\delta_M M$

представляють джерело та загибель макрофагів у здорових нормальних умовах, тоді як $\sum M$ враховується для пухлиногенної відповіді імунної системи, викликану інфікованими клітинами y . У рівнянні (4), перший член у правій частині - це виробництво TNF- α макрофагами, тоді як решта два члени є втрата шляхом поглинання у межах клітин і втрата через природну деградацію. Рівняння вірусу (5) містить вірус частинок із мертвих клітин y і втрати від поглинання x (тобто $\rho x v$) і природної деградація/кліренс ($\delta_v v$). У модель також включено неперервну ін'єкцію u_1 вірусу як вірусотерапевтичного препарату та безперервну ін'єкцію u_2 як TNF- α інгібітор (у рівнянні (4)).

Нехай $u_1 = \text{const} = C$, $u_2 = \text{const} = D$, але доцільно використовувати u_1 , u_2 як залежні від часу адаптивні керуючі змінні, щоб визначити оптимальні стратегії лікування гліоми. Оскільки розмірність густини і концентрації в (1)-(5) є $\tilde{n}/\tilde{m}^3$, $\frac{\rho}{\beta} = \frac{\text{virusmass}}{\text{viruscells}}$ є мале число. очікується, що b_1 буде дуже малим (віруси пошкоджуються під час некрозу), тому для простоти приймемо $b_1 = 0$. Значення b вибух (або реплікація) вірусів відіграватиме головну роль у прогресуванні захворювання та його лікуванні.

Позначимо через $n(t)$ густину усіх мертвих клітин. Тоді, на додаток до динаміки (1)-(5), маємо рівняння

$$\frac{dn}{dt} = k y \frac{T}{K + T} + \delta_y y + \delta_x x + \delta_M M - \mu n, \quad (6)$$

де μ – швидкість, з якою мертві клітини очищаються з пухлини.

Це рівняння буде враховуватися при розрахунку радіуса пухлини.

Табл. 9.1. Параметри моделі

Пар.	Опис	Значення	Розмірність
α	швидкість проліферації	0,2	1/доб
	неінфікованих пухлинних клітин		
β	інфікованість пухлин. клітин вірусами	$2 \cdot 10^4$	$\tilde{n}^3/(\tilde{a} \cdot \tilde{a} \cdot \tilde{a})$
ρ	швидкість втрати вірусів при зараженні	0,04	$\tilde{n}^3/(\tilde{a} \cdot \tilde{a} \cdot \tilde{a})$
k	ефективність інгібуючої дії TNF- α	0,4	1/доб
δ_y	швидкість загибелі	0,2	1/доба
	інфікованих пухлинних клітин		
λ	швидкість продукції TNF- α	$2,86 \cdot 10^{-3}$	1/доба
δ_T	швидкість деградації клітин TNF - α	55,45	1/доба
δ_M	смертність макрофагів	0,015	1/доба
b	розмір вибуху інфікованих клітин	(50 ÷ 150)	
	під час апоптозу	$\cdot 10^{-6}$	
b_1	розмір вибуху інфікованих клітин	$0 \ll b$	
	під час некрозу		
K	несуча здатність TNF- α	$5 \cdot 10^{-7}$	г/см ³
κ	деградація TNF- α	$4 \cdot 10^{-10}$	1/доба
	завдяки дії на інфіковані клітини		
δ_v	швидкість лізису вірусу	0,5	1/доб

A	постійне джерело макрофагів	$9 \cdot 10^{-7}$	г/($\tilde{n}^3$ доб)
s	швидкість стимуляції макрофагів	0,15	1/доб
	інфікованими клітинами без стимулу		
δ_x	рівень смертності	0,1	1/доб
	неінфікованих ракових клітин		
μ	швидкість видалення мертвих клітин	0,25	1/доб
θ_0	середнє загальної густини клітин	0,9	г/см ³
C	постійна інфузія вірусу	$(0 \div 5) \cdot 10^{-7}$	г/($\tilde{n}^3$ доб)
D	постійна інфузія інгібітору TNF- α	$0 \div 30$	

Розрахунок радіусу пухлини

Припускаємо, що пухлина сферична зі змінним радіусом $R(t)$, об'ємом $V(t)$ і масою

$$m(t) = x(t) + y(t) + M(t) + n(t).$$

У кожній точці кулі густина клітин зростає зі швидкістю dm/dt . Додаванням рівнянь (1)-(3) і (6) ми отримуємо, що

$$\frac{dm}{dt} = \alpha x - \delta_x x + A + syM - \mu n.$$

Загальна маса пухлини зростає зі швидкістю

$$V(t)(\alpha \tilde{x} - \delta_x \tilde{x} + A + syM - \mu \tilde{n}),$$

де $\tilde{(\cdot)}$ позначає середнє значення. Нехай θ_0 є сума середніх, $\theta_0 = \tilde{x} + \tilde{y} + \tilde{M} + \tilde{n}$. Ми припускаємо, що збільшення загальної маси призводить до збільшення об'єму пухлини пропорційно, тобто $\theta_0 dV/dt$ і що $yM = \tilde{y}\tilde{M}$, так що

$$\theta_0 \frac{dV}{dt} = V(t)(\alpha \tilde{x} - \delta_x \tilde{x} + A + syM - \mu \tilde{n}).$$

Оскільки

$$\frac{1}{V} \frac{dV}{dt} = \frac{3}{R} \frac{dR}{dt},$$

отримуємо

$$\theta_0 \frac{3}{R} \frac{dR}{dt} = (\alpha \tilde{x} - \delta_x \tilde{x} + A + s\tilde{y}\tilde{M}) - \mu \tilde{n}.$$

Отже

$$\theta_0 \frac{3}{R} \frac{dR}{dt} = (\alpha \tilde{x} - \delta_x \tilde{x} + A + s\tilde{y}\tilde{M}) - \mu(\theta_0 - \tilde{x} - \tilde{y} - \tilde{M}).$$

Нарешті, для спрощення розрахунків припустимо, що середні $\tilde{x}, \tilde{y}, \tilde{M}$ і $\tilde{v}$ задовольняють ті самі рівняння, що й x, y, M, v . Отже, ми отримуємо таку формулу для радіусу пухлини:

$$\theta_0 \frac{3}{R} \frac{dR}{dt} = (\alpha x - \delta_x x + A + syM) - \mu(\theta_0 - x - y - M). \quad (7)$$

Визначення параметрів

У табл. 8.1 наведено значення параметрів, які будуть використані в нашому аналізі. Пояснення щодо розрахунків значень параметрів наведено в додатку.

Оцінка параметрів

$$\alpha = 0,2 \text{ 1/доба}, \delta_x = 0,1 \text{ 1/доба}, \mu = 0,2 \text{ 1/доба}.$$

Нехай інфіковані клітини гинуть швидше за x клітин, і приймаємо $\delta_y = 0,2 \text{ 1/доба}$.
Діаметр вірусу $0,13 \mu m$. Діаметр клітини 10 мкм . Звідси маса $5 \cdot 10^5$ вірусів приблизно дорівнює масі однієї клітини. Маса клітини - 10^{-9} г . Отже, маса одного вірусу $2 \cdot 10^{-15} \text{ г}$. З [R5, R8] ми отримуємо

$$7 \cdot 10^{-10} < \beta < 1,7 \cdot 10^{-8} \text{ mm}^3/(\text{virus} \cdot \text{доба}).$$

Ми беремо

$$\beta = 10^{-9} \text{ mm}^3/(\text{virus} \cdot \text{доба}),$$

що в одиницях $\text{см}^3/(\text{г} \cdot \text{доба})$ дає

$$\beta = 10^{-8} \frac{10^{-3}}{2 \cdot 10^{-5} \text{ г.доба}} \frac{\text{см}^3}{\text{г.доба}} = 2 \cdot 10^4 \frac{\text{см}^3}{\text{г.доба}}.$$

Параметр ρ виражається в одиницях $\frac{\text{см}^3}{\text{г.доба}}$. Отже

$$\rho = \frac{\beta}{5 \cdot 10^5} = 4 \cdot 10^{-2} \frac{\text{см}^3}{\text{г.доба}}.$$

Значення $K = 5 \cdot 10^{-7} \text{ г/см}^3$ взято з [R17] і значення $\delta_T = 55,45 \text{ 1/доба}$ взято з [R9].
Ми припускаємо, що якщо $T + K$, рівень смертності y від T подвоює смертність δ_y , що дає $k = 0,4 \text{ 1/доба}$. Відношення $\kappa/k = 10^{-9}$. Отже,

$$\kappa = 10^{-9} k = 0,4 \cdot 10^{-9} \text{ 1/доба}.$$

У [R9] рівень смертності макрофагів залежить від їх фенотипу: для фенотипу M_1 рівень смертності становить $0,02 \text{ 1/доба}$, а для фенотипу M_2 — $0,008/\text{доба}$. Відповідно, $\delta_M = 0,015 \text{ 1/доба}$. Джерелом активованих макрофагів є здоровий мозок конститутивно малий; беремо $a = 9 \cdot 10^{-7} \text{ г/(\см}^3 \text{ доба)}$. Параметр s важко визначити; ми беремо $s = 0,15 \text{ г/(\см}^3 \text{ доба)}$. $\lambda = 2,86 \cdot 10^{-3}/\text{доба}$.

9.2.2 Результати

Моделюємо (1) - (7), щоб визначити, як стан системи реагує на комбіновану терапію. Значення параметрів для моделювання взято з табл. 8.1, якщо не зазначено інше. Ми припускаємо, що процес починається з таких початкових умов:

$$x(0) = 0,7, y(0) = 0,1, n(0) = 0, M(0) = 0,1, T(0) = 10^{-7}, 4/3\pi R(0)^3 = 0,9,$$

Поведінка моделі без терапії (некерована система)

Припустимо, що спочатку введено дозу вірусів, задану $v(0) = 10^{-6}$, і ми починаємо шляхом вивчення поведінки системи, коли не проводиться додаткова терапія.

На рис. 8.3 показано профіль усіх змінних моделі за перші 50 днів для різних

значень пакетного числа b . Якщо b невелике, а саме $b = 70$, радіус пухлини збільшується від початкового значення $R(0) = 0,6$ до 43 при $T = 50$, тобто $R(50) = 43$ см. Для $b = 90$ $R(50) = 6$ см. Лише коли $b = 150$ $R(50)$ більше не збільшується відносно $R(0)$.

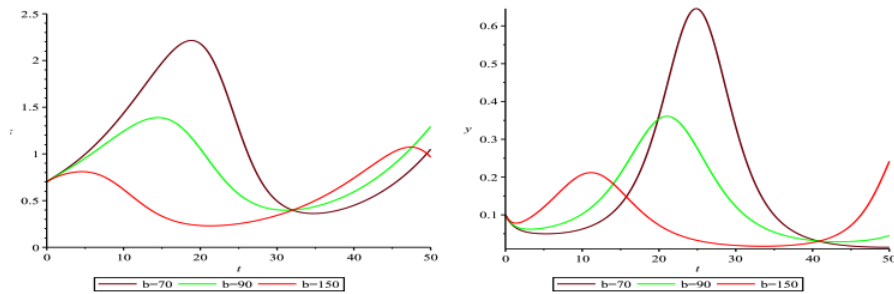


Рис. 9.3 Графіки змінних моделі для $C = 0, D = 0$

Моделювання $T(t)$ з $b = 70$ свідчить, що початкове навантаження $v(0)$ призводить до значного підвищення $TNF-\alpha$, яке пізніше нормалізується. Після початкової ін'єкції вірусу вірус розмножується, щоб досягти свого максимуму на 28 день; після цього часу густина вірусу продовжує зменшуватися – препарат “занадто слабкий”. В результаті зростання R після 30-го дня обмежено. Проте зростаюча кількість вірусу у перші 30 днів спричиняє зростання M , що призводить до зростання T . І, як і раніше, великі кількості T стримують реплікацію вірусу, що міститься в інфікованих клітинах y – що погано для терапії.

Переходячи до випадку з $b = 90$, показаного на рис. 8.3, ми помітимо, що поведінка T майже така сама як і в попередньому випадку. Радіус R збільшується з часом, але не так різко, як раніше. Нарешті, у випадку $b = 150$, як показано на рис. 8.3, зростання v відбувається швидше, але так само воно потім зменшується, щоб нарешті відновити зростання. Ми також бачимо, що $R(t)$ спочатку трохи зростає, потім зменшується і, нарешті, знову трохи збільшується, імітуючи коливання v .

Ця поведінка засвідчує, що якщо ми хочемо досягти зниження розміру пухлини без неперервної терапії, а лише з початковою вірусною ін'єкцією $t = 0$, потрібно буде ввести дуже велику (імовірно, нереалістичну) кількість вірусу. Це говорить про те, що необхідна неперервна терапія ін'єкцій вірусу.

Поведінка моделі з неперервною постійною інфузією вірусу

Дослідимо поведінку системи за припущення, що застосовується постійна ін'єкція вірусу $u_1(t) \equiv C = 5 \cdot 10^{-7}$ і початкове навантаження $v(0) = C$, знову для трьох вірусних значень реплікації, $b = 70$, $b = 90$ і $b = 150$. На рис. 8.4 показано еволюцію системи за цих припущень. Ми бачимо таку ж коливальну поведінку, як і раніше.

Але, що важливіше, ми бачимо вплив препарату на радіус пухлини: для $b = 70$ і $b = 90$ радіус $R = R(t)$ все ще зростає, хоча набагато менше за випадок $C = 0$ (неперервної інфузії). Але для кількості пакетів $b = 150$ радіус пухлини, після невеликого початкового збільшення, строго зменшується, з $R(50)$ приблизно до половини початкового радіуса пухлини.

У випадку $b = 70$ радіус $R(t)$ збільшується, але розмір через 50 днів майже у 5 разів менший за випадок на рис. 8.3, хоча ми дали меншу початкову дозу на випадок без

терапії. Знову $v(t)$ досягає максимального значення (тепер на 25 день), але воно потім не вмирає. Крім того, зростання M і, отже, T , менше, і це має менший вплив на вбивство $y(t)$. Переходячи до випадку $b = 90$, ми все ще не у найбажанішій ситуації - радіус все ще зростає, хоча у 3 рази менше за випадок на рис. 8.3 на 50-й день. Як і можна було очікувати, для $b = 150$ радіус пухлини зменшується, але тільки після короткочасного збільшення. Це початкове збільшення, ймовірно, через те, що початкове вірусне навантаження невелике.

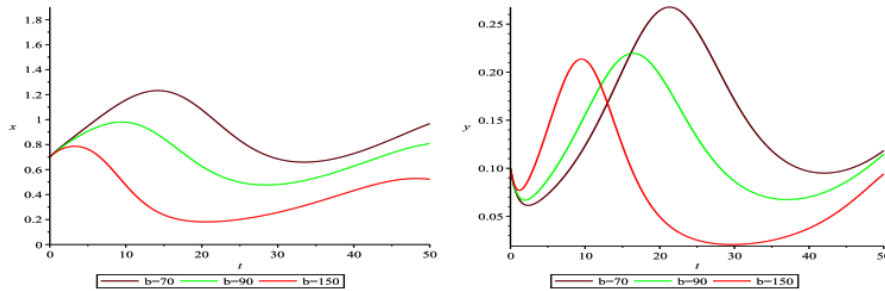


Рис. 9.4 Графіки змінних моделі для $C = 5 \cdot 10^{-7}$, $D = 0$

Поведінка моделі при комбінованій терапії

Далі будемо вважати, що комбінована терапія проводиться з постійними темпами, тобто $u_1 \equiv C$, $u_2 \equiv D$. Покладемо $C = 2 \cdot 10^{-7}$, $D \equiv 15$. Знову ми вивчимо відповідь системи на різні рівні виробництва вірусу інфікованими клітинами.

Відповіді для $b = [70; 90; 150]$ показані на рис. 8.5, який аналогічний до рис. 8.4.

По-перше, при $b = 70$ розмір пухлини збільшується, але не так сильно, як у випадку $D = 0$. Графік $T(t)$ виглядає подібно до випадку без D -терапії, значення $T(t)$ на порядок менші за попередні – зараз максимум близький до $3,3 \cdot 10^{-7}$ порівняно із $7 \cdot 10^{-6}$, де $D = 0$. Застосування u_2 також підтримує ріст v - тепер приблизно у 5 разів вище. Цікаво, це хоча T зменшується, M більше за попереднє.

Для $b = 90$ радіус пухлини, після початкового збільшення, знижується, щоб досягти початкового розміру. Нарешті, для $b = 150$ отримуємо бажану ціль щодо зменшення пухлини; радіус значно зменшується приблизно до $R(50) = 0,1$ см.

Тепер стає природним придивитися ближче до самої терапії та побачити, як обидва основні агенти, інфузія вірусу та інгібітор $TNF-\alpha$, сприяють успіху терапії. Іншими словами, який внесок кожного з них для зменшення радіусу пухлини.

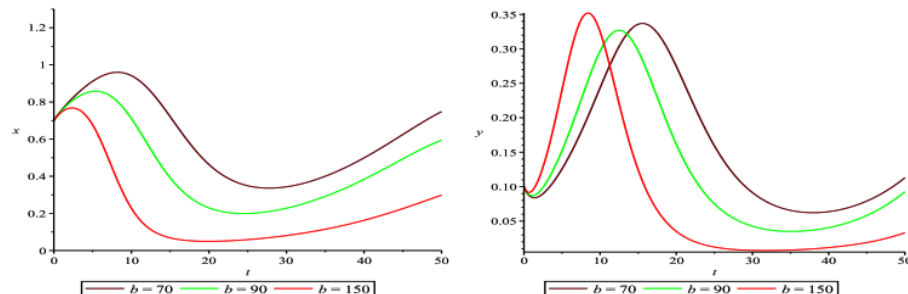


Рис.9.5 Графіки модельних змінних $C = 5 \cdot 10^{-7}$, $D = 15$

Видно, що навіть при $b = 80$ можна зменшити $R(50)$ з $R(0)$, якщо взяти $\tau = 30$,

тобто якщо ми обмежимо продукцію $\text{TNF-}\alpha$ на 97%.

Підводячи підсумок, мета щодо зменшення розміру пухлини може бути досягнута різними способами, тому, вибираючи найкращу комбінацію, слід враховувати можливі серйозні побічні ефекти лікування. Це підводить нас до основної теми цієї роботи, яка полягає в оцінці ефективності обох методів лікування.

9.3 Вплив CNN1 і макрофагів на гліому

У цьому розділі розглядається математична модель, що є розвиненням моделі [J7].

Табл.9.2 Змінні моделі

Пар.	Опис	Розмірність
x	густина неінфікованих пухлинних клітин	клітин/ мм^3
y	густина інфікованих пухлинних клітин	клітин/ мм^3
n	густина некротичних пухлинних клітин	клітин/ мм^3
M	густина клітин макрофагів	клітин/ мм^3
v	густина вільних вірусів	вірусів/ мм^3
P	Концентрація MCP-1	$\text{г}/\text{мм}^3$
C	Концентрація CCN1	$\text{г}/\text{мм}^3$
u	Швидкість адвекції	$\text{мм}/\text{добу}$
R	Радіус пухлини	мм

9.3.1 Математична модель

Математична модель базується на мережі, показаній на рис. 8.2. Змінні моделі наведені в табл. 8.2. Описи параметрів моделі наведено в табл. 8.3.

Пухлинні клітини (x) інфікуються (y) позаклітинним вірусом (v). При лізисі інфіковані клітини стають мертвими (n), вивільняючи свої новоутворені віруси до ЕСМ.

Табл.9.3 Параметри моделі

Парам.	Опис
D	коефіцієнт дифузії клітин
D_v	Коефіцієнт дифузії вільних вірусів
D_P	коефіцієнт дифузії MCP-1
D_C	Коефіцієнт дифузії CCN1
	хемотаксична чутливість макрофага
χ_P	до MCP-1
χ_C	до CCN1
λ_x	швидкість проліферації
	неінфікованих пухлинних клітин
λ_M	активація макрофагів
λ_P	Швидкість виробництва MCP-1
λ_C	швидкість виробництва CCN1

σ_P	джерело MCP-1
σ_C	джерело CCN1
k_x	CCN1 інгібування проліферації
	неінфікованих клітин
k_y	CCN1 інгібування лізису

Табл.9.4 Параметри моделі

Парам.	Опис
d_y	швидкість лізису інфікованих п. кл.
d_n	швидкість видалення мертвих п. кл.
d_M	смертність макрофагів
d_v	рівень очищення від вільних вірусів
d_P	швидкість деградації MCP-1
d_C	швидкість деградації CCN1
	швидкість знищення макрофагами
δ_y	інфікованих пухлинних клітин
δ_v	вільних вірусів
α_C	знищення макрофагів, посилені CCN1
α_{MC}	активація макрофагів посилена CCN1
α_M	продукція MCP-1, посилена макрофагами
k_C	насичення CCN1 для знищення макрофагів
k_{MC}	насичення CCN1 для активації макрофагів
k_M	насичення макрофагів для виробництва MCP-1
β_x	рівень інфікування неінфікованих
	пухлинних клітин вільними вірусами
β_v	швидкість поглинання вільних вірусів
	неінфікованими пухлинними клітинами
b	розмір вірусного вибуху лізованої
	інфікованої пухлинної клітини
θ	густина клітин
$\tilde{\alpha}$	макрофагальна інфільтрація, індукована MCP-1
k_P	насичення MCP-1 для інфільтрації макрофагами
$\tilde{M}$	густина макрофагів у крові

Інфіковані пухлинні клітини виділяють білок CCN1, і макрофаги притягуються до пухлини за допомогою CCN1 і MCP-1. Наявність MCP-1 у мікрооточенні пухлини виявлена у [J7]. Ми припускаємо, що MCP-1 виробляється інфікованими клітинами і що виробництво посилюється макрофагами.

Ми переходимо до розробки динамічної системи на основі мережі, рис. 8.2.

Система диференціальних рівнянь включає дифузію для клітин, вірусів і білків, а також адвекцію клітин з однаковою швидкістю u для всіх клітин. Тому для

обчислювальних цілей ми припускаємо, що пухлина є сферична, а радіус росте з часом: $R(t)$. Ми додатково припускаємо, що всі змінні є сферично симетричні, так що $u = u(r, t)z/r$, де z є точка у 3d просторі R^3 з $|z| = r$.

Неінфіковані пухлинні клітини (x). Проліферація неінфікованих пухлинних клітин відбувається зі швидкістю λ_x , і ми припускаємо, що проліферація пригнічується за присутності CCN1 через регуляцію медіаторів відповіді інтерферону I типу (ІФН). Клітини стають заражені вірусом зі швидкістю, пропорційною β , яка враховує ймовірність успішного проникнення вірусу при контакті з клітиною. Рівняння для густини неінфікованої пухлинної клітини $x(r, t)$ визначається як

$$\frac{\partial x}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u x) = D \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial x}{\partial r} \right) + \frac{\lambda_x}{1 + k_x C} x - \beta_x x v. \quad (8)$$

Другий член у лівій частині рівняння описує адвекцію, перший член праворуч описує дифузію, другий - проліферацію, третій -- інфекцію. Джерелом дифузії (або розсіювання) клітин є безладний рух частинок, які мають тенденцію переходити від високих концентрацій до низьких.

Інфіковані пухлинні клітини (y). Інфіковані пухлинні клітини піддаються лізису зі швидкістю d_y через високе розмноження вірусу, що викликає розрив клітин. Лізис пригнічується CCN1 активацією клітинної противірусної захисної відповіді. Інфіковані клітини також можуть бути вбиті макрофагами. Отже, еволюція густини інфікованої пухлинної клітини $y(r, t)$ визначається рівнянням

$$\frac{\partial y}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u y) = D \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial y}{\partial r} \right) = \beta_x x v - \delta_y y M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - \frac{d_y}{1 + k_y C} y. \quad (9)$$

Третій член праворуч описує вбивство макрофагами, 4-й -- лізис.

Мертві пухлинні клітини (n). Ми враховуємо наявність клітин, які були вбиті макрофагами або лізисом шляхом включення густини мертвих пухлинних клітин, $n(r, t)$. Припустимо, що ці мертві клітини видаляються зі швидкістю d_n . Тоді рівняння для $n(r, t)$ має вигляд

$$\frac{\partial n}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u n) = D \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial n}{\partial r} \right) + \delta_y y M \left(1 + \frac{\alpha_C C}{k_C + C} \right) + \frac{d_y}{1 + k_y C} y - d_n n. \quad (10)$$

Макрофаги (M). Густина прозапальних макрофагів $M(r, t)$ визначається за рівнянням

$$\begin{aligned} \frac{\partial M}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u M) = D \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial M}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \chi_P M \frac{\partial P}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \chi_C M \frac{\partial C}{\partial r} \right) \\ + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M. \end{aligned} \quad (11)$$

Другий і третій члени у рівнянні праворуч описують хемотаксис MCP-1 і CCN1, 4-й -- активацію, 5-й -- смерть.

На додаток до адвекції та дифузії макрофаги зазнають спрямованого руху через хемотаксис. Хоча MCP-1 є відомим хемоаттрактантом для макрофагів, нещодавно отримані дані показують, що CCN1 також може індукувати пряму міграцію в макрофагах, [J7].

Клітини гліобластоми виділяють різні цитокіни, які приваблюють асоційовані з

пухлиною макрофаги (ТАМ). Зразки біопсії GBM показали проникнення та локалізацію ТАМ всередині пухлини без різниці між центральною та периферичною областями пухлини. Тому для простоти вибираємо джерело макрофагів, $\lambda_M = \text{const}$.

Вільний вірус (v). Густина вільного вірусу моделюється згідно з рівнянням

$$\frac{\partial v}{\partial t} = D_v \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial v}{\partial r} \right) + \frac{bd_y}{1+k_C C} y - \delta_v v M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - \beta_v x v - d_v v. \quad (12)$$

Вільні віруси вивільняються шляхом лізису інфікованих клітин розміром спалаху b . Ми припускаємо, що макрофаги поглинають вільний вірус зі швидкістю δ_v , а CCN1 ще більше посилює це вбивство, як це відбувається з інфікованими клітинами. Вільний вірус втрачається через проникнення вірусу у пухлинну клітину (при інфікуванні), а також через механізми очищення вірусу.

MCP-1 (P). Моноцитний хемотаксичний білок-1 (MCP-1), який також зазвичай називають CCL2 виробляється інфікованими клітинами гліоми, [J7]. Припускається, що також є постійне джерело MCP-1, σ_P , але явно не включає експресію MCP-1 здоровими клітинами гліоми. Еволюція концентрації MCP-1, $P(r, t)$, визначається рівнянням

$$\frac{\partial P}{\partial t} = D_P \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial P}{\partial r} \right) + \sigma_P + \lambda_P y \left(1 + \frac{\alpha_M M}{k_M + M} \right) - d_P P, \quad (13)$$

третій член праворуч (виробництво) враховує перехресні перешкоди між макрофагами і клітинами гліоми.

CCN1 (C). CCN1 виробляється головним чином пухлинними клітинами всередині пухлинної маси, що росте. Результати [J8], рис. 8.2 вказують на те, що, хоча є деякі CCN1 гени експресії у клітинах гліоми за відсутності онколітичного вірусу, експресія значно підвищується після інфікування онколітичним вірусом. Тому припустимо, що існує постійне виробництво CCN1, σ_C , а також виробництво CCN1 шляхом інфікованих клітин зі швидкістю λ_C . Рівняння, що визначає динаміку концентрації CCN1, визначається як

$$\frac{\partial C}{\partial t} = D_C \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial C}{\partial r} \right) + \sigma_C + \lambda_C y - d_C C. \quad (14)$$

Швидкість адвекції (u). Ми моделюємо пухлину як нестисливу рідину, в якій поле швидкості виникає внаслідок проліферації та видалення пухлинних клітин. Вважаємо, що загальна густина макрофагів і пухлинних клітин постійна, позначимо θ , тобто

$$x + y + n + M = \theta \quad (15)$$

по всій пухлині. Підсумовування рівнянь (8)–(11) дає

$$\begin{aligned} \frac{\theta}{r^2} \frac{\partial}{\partial r} (r^2 u) = & \frac{\lambda_x}{1+k_x C} - d_n n - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \chi_P M \frac{\partial P}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \chi_C M \frac{\partial C}{\partial r} \right) \\ & + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M. \end{aligned} \quad (16)$$

і шляхом інтегрування ми отримуємо таке рівняння для u , швидкості адвекції:

$$u = \frac{1}{\theta r^2} \int_0^r s^2 \left[\frac{\lambda_x}{1+k_x C} x - d_n n + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - \frac{d_M}{M} \right] ds$$

$$-\frac{\chi_p}{\theta} M \frac{\partial P}{\partial r} - \frac{\chi_c}{\theta} M \frac{\partial C}{\partial r}. \quad (17)$$

Межові умови

Будемо вважати, що радіус пухлини змінюється з локальною швидкістю, тобто

$$\frac{dR}{dt} = u, \quad r = R(t). \quad (18)$$

Межові умови у центрі пухлини випливають із передбачуваної сферичної симетрії:

$$\frac{\partial x}{\partial r} = \frac{\partial y}{\partial r} = \frac{\partial n}{\partial r} = \frac{\partial M}{\partial r} = \frac{\partial v}{\partial r} = \frac{\partial P}{\partial r} = \frac{\partial C}{\partial r} = 0, \quad r = 0. \quad (19)$$

Ми приймаємо межову умову відсутності потоку для густини вільного вірусу на межі пухлини,

$$\frac{\partial v}{\partial r} = 0, \quad r = R(t). \quad (20)$$

Ми припускаємо, що пухлина знаходиться в передметастатичному стані, отже, ракові клітини не видаляються з пухлини. Оскільки білки MСР-1 і ССN1 виробляються пухлинними клітинами, ми далі припускаємо, що білки, як і пухлинні клітини, слідуєть межовій умові без потоку. Ці припущення призводять до таких межових умов для густини пухлинних клітин і концентрації білка на межі пухлини:

$$\frac{\partial x}{\partial r} = \frac{\partial y}{\partial r} = \frac{\partial n}{\partial r} = \frac{\partial P}{\partial r} = \frac{\partial C}{\partial r} = 0, \quad r = R(t). \quad (21)$$

Отже, ми припускаємо, що макрофаги у кровоносних судинах поза пухлиною, де густина макрофагів дорівнює $\tilde{M}$, інфільтрують пухлину внаслідок залучення МСР-1. Маємо:

$$\frac{\partial M}{\partial r} + \alpha(M - \tilde{M}) = 0, \quad r = R(t). \quad (22)$$

з функціональною формою

$$\alpha(P) = \tilde{\alpha} \frac{P}{k_p + P};$$

ми не включили залежність від C , припускаючи, що МСР-1 є сильнішим за ССN1 атрактантом моноцитів із крові.

9.3.2 Оцінка параметрів і вміст макрофагів

Повна інформація про оцінку параметрів для моделі наведена в Додатку 1. Ми виділяємо тут оцінку параметрів, на які впливає вміст макрофагів пухлини, параметр, який показано в розд. 4, відіграє значну роль у визначенні результатів лікування.

Позначимо m ($m = 0,10$) -- вміст макрофагів у пухлині, тобто співвідношення густини макрофагів до загальної густини клітин у пухлині. Густина макрофагів у пухлині, M_0 , задана як

$$M_0 = m\theta, \quad (23)$$

де θ – загальна густина пухлинних клітин і макрофагів (рівняння (15)). Дослідження зразків гліобластоми людини вказують на вміст макрофагів у діапазоні 1,5–15% [9]. Усі параметри моделі та концентрації, що використовуються у випадку за замовчуванням, як оцінено у додатку 1, можна знайти в табл. 8.5 і 8.7 відповідно.

Далі досліджується моделювання зі змінними значеннями m від 0,01 до 0,3. Параметри для активації макрофагів, λ_m , і насичення макрофагів для виробництва МСР-1, k_m , є параметри моделі, на які впливає зміна значення m . Як λ_m , так і k_m зростають лінійно з m .

9.3.3 Числові результати

Спочатку ми зменшуємо розміри моделі та робимо перетворення, щоб зафіксувати рух межі, як описано в Додатку R.

Початковий профіль вірусу $v_0(r) = v(r, 0)$ із центрованим розподілом Гауса при $r = 0$:

$$v_0(r) = a \exp\left(\frac{-r^2}{2\sigma^2}\right), \quad (24)$$

де $a = 2,5$, $\sigma = 0,48$.

Інші початкові умови задані формулою

$$x_0(r) = 1 - \tilde{M}, y_0(r) = 0, M_0(r) = \tilde{M}, P_0(r) = 1, C_0(r) = 1, R(0) = 3 \text{ mm}; \quad (25)$$

тобто ми припускаємо, що спочатку немає інфікованих або мертвих пухлинних клітин.

Табл.9.5 Значення параметрів моделі

Парам.	Значення	Безрозмірне значення
D	$8,64 \cdot 10^{-5} \text{ mm}^2/\text{доба}$	-
D_v	$0,864 \text{ mm}^2/\text{доба}$	-
D_p	$17,3 \text{ mm}^2/\text{доба}$	-
D_C	$392,9 \text{ mm}^2/\text{доба}$	-
χ_P	$10^6 \text{ mm}^5/(\text{г.доба})$	$3 \cdot 10^{-7} \text{ mm}^2/\text{доба}$
χ_C	$10^6 \text{ mm}^5/(\text{г.доба})$	$10^{-3} \text{ mm}^2/\text{доба}$
λ_x	$0,2 \text{ 1/доба}$	-
λ_m	$1,07 \cdot 10^4 \text{ кл./}(\text{мм}^3 \text{ доба})$	$1,07 \cdot 10^{-2} \text{ 1/доба}$
λ_P	$7,65 \cdot 10^{-18} \text{ г/}(\text{кл. доба})$	$25,5/\text{доба}$
λ_C	$1,1 \cdot 10^{-14} \text{ г/}(\text{кл. доба})$	$11/\text{доба}$
σ_P	$5,1 \cdot 10^{-13} \text{ г/}(\text{мм}^3 \text{ доба})$	$1,7 \text{ 1/доба}$
σ_C	$5,5 \cdot 10^{-10} \text{ г/}(\text{мм}^3 \text{ доба})$	$0,55 \text{ 1/доба}$
k_x	$10^7 \text{ мм}^3/\text{г}$	$0,01$
k_y	$10^8 \text{ мм}^3/\text{г}$	$0,1$
d_y	$1,5 \text{ 1/доба}$	-
d_n	$0,5 \text{ 1/доба}$	-
d_M	$0,015 \text{ 1/доба}$	-

d_v	0,5 1/доба	-
d_P	1,7 1/доба	-
d_C	0,55 1/доба	-
δ_y	$4,8 \cdot 10^{-8}$ мм ³ /(кл. доба)	0,048 1/доба
δ_v	$1,2 \cdot 10^{-13}$ мм ³ /(кл. доба)	$1,2 \cdot 10^{-6}$ 1/доба

Табл. 9.6 Значення параметрів моделі

Парам.	Значення	Безрозмірне значення
α_C	1	-
α_{MC}	2	-
α_M	11,7	-
k_C	$2 \cdot 10^{-9}$ г/мм ³	2
k_{MC}	10^{-9} г/мм ³	1
k_M	10^5 кл./мм ³	0,1
β_x	$1,7 \cdot 10^{-8}$ мм ³ /(вірус доба)	0,88 1/доба
β_v	$1,7 \cdot 10^{-8}$ мм ³ /(вірус доба)	0,017 1/доба
b	50 вірус/доба	-
θ	10^6 кл./мм ³	-
$\tilde{\alpha}$	1	-
k_P	$3 \cdot 10^{-13}$ г/мм ³	1
$\tilde{M}$	50 кл./мм ³	$5 \cdot 10^{-5}$

Безрозмірне значення «-» означає, що параметр не змінився

Табл. 9.7. Концентрація білка та густина клітин

Позн.	Опис	Значення
C_0	типова концентрація CNN1	10^{-9} г/мм ³
C_1	концентрація CNN1 при OV лікуванні	$3 \cdot 10^{-9}$ г/мм ³
M_0	типова густина макрофагів у пухлині	10^5 кл./мм ³
M_1	густина макрофагів при OV лікуванні	$1,5 \cdot 10^5$ кл./мм ³
P_0	типова концентрація MCP-1	$3 \cdot 10^{-13}$ г/мм ³
P_1	концентрація MCP-1	
	(введення без макрофагів)	$7,5 \cdot 10^{-13}$ г/мм ³
P_2	концентрація MCP-1	
	(введення, макрофаги)	$3,9 \cdot 10^{-12}$ г/мм ³
Y_0	апрокс. густини	
	введеної пухлинної клітини	10^5 кл./мм ³

9.3.4 Додаток

Оцінка параметрів для розмірної моделі

Коефіцієнти дифузії та хемотаксису

Вважаємо, що коефіцієнти дифузії всіх пухлинних та імунних клітин рівні і прийняти їх рівними $8,64 \cdot 10^{-5} \text{ мм}^2/\text{доба}$, [10]. Ми припускаємо, що коефіцієнт дифузії D_p для білка P можна розрахувати за його молекулярною масою M_p за формулою $D_p = K \sqrt[3]{M_p}$, де $K = 0,325 \text{ мм}^2/(\text{доба} \cdot \sqrt[3]{\text{dalton}})$. Молекулярна маса CCN1 становить 42027 дальтон. Отже, коефіцієнт дифузії CCN1 оцінюється за

$$D_C = K \sqrt[3]{M_{CCN1}} = 0,325 \cdot \sqrt[3]{42027} = 392,9 \text{ мік}^2/\text{доба}.$$

Чутливість макрофагів до хемоаттрактантів важко оцінити і визначити в літературі відносять діапазон від $8,64 \cdot 10^4 \div 1,73 \cdot 10^9 \text{ мм}^5/(\text{г. доба})$, [J10]. Вибираємо проміжне значення

$$\chi_C = \chi_P = 10^6 \text{ мм}^5/(\text{г. доба}).$$

Оцінка d_C, λ_C, μ_C із рівняння (14).

Припускається, що період піврозпаду позаклітинного CCN1 становить 30 годин і відповідно

$$d_C = \frac{\ln 2}{30} 1/200 = 0,551/\text{доба}.$$

CCN1 має високу надмірну експресію в гліомі, тому ми розглядаємо типову концентрацію CCN1 у пухлині $C_0 = 10^{-9} \text{ г/мм}^3$. Стаціонарний стан

$$\sigma_C + \lambda_C y - d_C C = 0. \quad (26)$$

Ми припускаємо, що за відсутності вірусу CCN1 досягає стаціонарної концентрації C_0 ; тому беремо

$$\sigma_C = d_C C_0 = 5,5 \cdot 10^{-10} \text{ г/мм}^3/\text{доба}.$$

Крім того, ми приймаємо густину пухлинних клітин і макрофагів (рівняння (15) як $\theta = 10^6 \text{ кл./мм}^3$ і припустимо, що приблизно одна десята пухлинних клітин інфікована вірусом. Отже, ми беремо типову концентрацію інфікованих пухлинних клітин $y_0 = 10^5 \text{ кл./мм}^3$. Експериментальні дослідження свідчать, що за наявності онколітичного вірусу експресія гена CCN1 порушується приблизно в три рази вище за відсутності вірусу. Відповідно, покладемо $C_1 = 3C_0$ і з рівняння (26) виводимо, що

$$\lambda_C = \frac{d_C C_1 - \sigma_C}{y_0} = 1,1 \cdot 10^{-14} \text{ г./ккл. доба}.$$

Оцінка $b, d_y, k_y, d_v, \delta_y, \delta_v$ за рівняннями (9) і (12). Цикл лізису приблизно 12 годин. Маємо оцінки коефіцієнтів:

$$k_y = 0,1/C_0 = 10^8 \text{ мм}^3/\text{г}; \quad b = 50 \text{ вірус / ккл}; \quad \delta_v = 0,5/\text{доба}.$$

$$\delta_y = 4,8 \cdot 10^{-8} \text{ мм}^3/\text{ккл. доба}.$$

Ми робимо припущення, що швидкість знищення, опосередкована макрофагами, пропорційна площі поверхні мішені. Припустимо, що пухлинні клітини мають сферичну форму з діаметром 40 мкм. HSV-1 є ікосаедричним вірусом з діаметром приблизно 200 нм. Отже, відношення площі поверхні пухлинної клітини до площі поверхні вірусу дорівнює

приблизно $4 \cdot 10^4$, і ми відповідно беремо

$$\delta_v = \frac{\delta_y}{4 \cdot 10^4} = 1,2 \cdot 10^{-12} \text{ мм}^3/(\text{кл. доба}).$$

Оцінка α_c , k_c за рівняннями (9) і (12) і m .

Приймаються значення за замовчуванням вмісту макрофагів, $m = 0,10$ і, згідно з рівнянням (23) густина макрофагів у типовій пухлині $M_0 = 10^5$ кл./мм³. Припускається, що густина макрофагів при лікуванні OV, M_1 , у 1,5 рази вище за типову пухлину. Оцінка члену в (9) для знищення макрофагами інфікованих клітин:

$$M_1 \delta_y y_1 \left(1 + \frac{\alpha_c C_1}{k_c + C_1} \right) = 1, 2 M_0 \delta_y y_0 \left(1 + \frac{\alpha_c C_0}{k_c + C_0} \right).$$

. Тут ми припускаємо, що кількість інфікованих клітин обернено пропорційна кількості макрофагів, так що $y_1 = y_0/1,5$, оскільки $M_1 = 1,5 M_0$. Прийнемо $k_c = 2 C_0$ і таким чином розв'яжемо для α_c , щоб отримати $\alpha_c = 1$.

Оцінка λ_x , k_x , β_x за рівнянням (8).

Оскільки висока експресія CCN1 корелює з поганим прогнозом, ми припускаємо, що інгібування росту пухлини через відповідь IFN, керовану CCN1, не є строга як інгібування вірусної реплікації у рівнянні (12). Тому оцінюємо $k_x = 0,01/C_0 = 10^7$ мм³/г, так що $k_x = 0,1 k_y$. Беремо швидкість розповсюдження неінфікованих клітин гліоми $\lambda_x = 0,2$ /доба на основі даних вимірювання росту підшкірних пухлин гліоми у контрольній групі мишей, які отримували фізіологічний розчин із фосфатним буфером, [J11]. Ми вибираємо рівень зараження клітин вірусом $\beta_x = 1,7 \cdot 10^{-8}$ мм³/(доба вірус).

Оцінка λ_m , α_{MC} , k_{MC} за рівнянням (11).

Експериментальні результати, [J4], свідчать, що OV-індукований CCN1 суттєво збільшує міграцію макрофагів і посилює прозапальну активацію макрофагів. Відповідно беремо $\alpha_{MC} = 2$. Також беремо $k_{MC} = C_0$. Швидкість загибелі макрофагів, $d_m = 0,015$ 1/доба, [J10].

Для оцінки постійного джерела макрофагів λ_m розглянемо динаміку густини макрофагів у просторово однорідному стаціонарному стані без вірусу. У цьому випадку існує невід'ємна швидкість адвекції, u , і рівняння (8) означає, що

$$\frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u) = \frac{\lambda_x}{1 + k_x C}.$$

Заміщення цього виразу в однорідний стаціонарний стан рівняння (11) дає

$$M \frac{\lambda_x}{1 + k_x C} = \lambda_m \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_m M,$$

так що

$$\lambda_m = \frac{\left(\frac{\lambda_x}{1 + k_x C_0} + d_m \right) M_0}{1 + \frac{\alpha_{MC} C_0}{k_{MC} + C_0}} = 1,07 \cdot 10^4 \text{ кл.мм}^3/\text{доба}. \quad (27)$$

Оцінка λ_p , α_m , k_m за рівнянням (13).

Ми припускаємо, що в нормальній здоровій тканині концентрація МСР-1 $P_0 = 3 \cdot 10^{-13}$ Г/мм³ і швидкість розпаду МСР-1 становить $d_p = 1,7$ 1/доба. Відповідно до рівняння (13) стаціонарна концентрація МСР-1 задовольняє рівняння

$$\sigma_p + \lambda_p y \left(1 + \frac{\alpha_M M}{k_M + M} \right) - d_p P = 0 \quad (28)$$

і ($y = 0$, $P = P_0$)

$$\sigma_p = d_p P_0 = 5,1 \cdot 10^{-13} \text{ г/(мм}^3 \text{ доба)}.$$

Експериментальні дані показують, що порівняно з неінфікованими клітинами гліоми експресія МСР-1 приблизно в 2,5 рази більше в інфікованих клітинах гліоми без макрофагів і приблизно в 13 разів вище в інфікованих клітинах гліоми, культивованих з макрофагами, [J4]. Відповідно, покладемо $P_1 = 2,5P_0$ і $P_2 = 13P_0$. За рівнянням (115) ми можемо оцінити λ_p , швидкість продукції МСР-1 інфікованими клітинами гліоми

$$\lambda_p = \frac{d_p P_1 - \sigma_p}{y_0} = 7,65 \cdot 10^{-18} \text{ г/(кл. доба)}$$

і α_M , що відповідає подальшій продукції МСР-1 через передачу сигналів макрофагами

$$\alpha_M = \frac{k_M + M_1}{M_1} \left(\frac{d_p P_2 - \sigma_p}{\lambda_p y_0} - 1 \right) = 11,7,$$

де ми це припустили, що $k_M = M_0$.

Оцінка $\tilde{\alpha}$, k_p , $\hat{M}$ із межових умов (22). Густина макрофагів у крові коливається від $2 \cdot 10^4$ до 10^5 клітин/мл. Відповідно виберемо $\hat{M} = 50$ кл./мм³. Також візьмемо $\tilde{\alpha} = 1$ і $k_p = P_0$.

Безрозмірна модель із фіксованою межею

Ми обезрозмірюємо густину клітин і вірусів за допомогою

$$\bar{x} = x/\theta, \bar{y} = y/\theta, \bar{n} = n/\theta, \bar{M} = M/\theta, \bar{v} = v/(b\theta).$$

Безрозмірні концентрації білка $\bar{P} = P/P_0$, $\bar{C} = C/C_0$. концентрації P_0, C_0 наведені в табл. 8.7. Відповідно інші безрозмірні параметри:

$$\begin{aligned} \bar{\beta}_x &= b\theta\beta_x, \bar{\beta}_y = \beta_y\theta, \bar{\delta}_y = \delta_y\theta, \bar{\delta}_v = \delta_v\theta, \\ \bar{\chi}_P &= \chi_P P_0, \bar{\chi}_C = \chi_C C_0, \bar{k}_x = k_x C_0, \bar{k}_y = k_y C_0, \bar{\lambda}_M = \lambda_M/\theta, \bar{\alpha}_{MC} = \alpha_{MC}/\theta, \\ \bar{k}_M &= k_M/\theta, \bar{M} = \tilde{M}/\theta, \bar{\sigma}_P = \sigma_P/P_0, \bar{\lambda}_P = \lambda_P\theta/P_0, \bar{k}_P = k_P/P_0, \\ \bar{\sigma}_C &= \sigma_C/C_0, \bar{\lambda}_C = \lambda_C\theta/C_0, \bar{k}_C = k_C/C_0, \bar{k}_{MC} = k_{MC}/C_0. \end{aligned}$$

Ми усуваємо рівняння для мертвих клітин за формулою (15) і зафіксуємо рухому межу перетворенням $\bar{r} = r/R(t)$. Відкинувши штрихові позначення для простоти, ми отримаємо наступну безрозмірну модель для $r \in (0,1)$, $t \geq 0$:

$$\frac{\partial x}{\partial t} = \frac{D}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial x}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{ux}{R} \right) + \frac{r}{R} \frac{dR}{dt} \frac{\partial x}{\partial r} + \frac{\lambda_x}{1+k_x C} x - \beta_x x v, \quad (29)$$

$$\begin{aligned}\frac{\partial y}{\partial t} = & \frac{D}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial y}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{uy}{R} \right) + \frac{r}{R} \frac{dR}{dt} \frac{\partial y}{\partial r} \\ & + \beta_x xv - \delta_y y M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - \frac{d_y}{1 + k_y C} y,\end{aligned}\quad (30)$$

$$\begin{aligned}\frac{\partial M}{\partial t} = & \frac{D}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial M}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \left(\frac{uM}{R} + \frac{\chi_P M}{R^2} \frac{\partial P}{\partial r} + \frac{\chi_C M}{R^2} \frac{\partial C}{\partial r} \right) \right] \\ & + \frac{r}{R} \frac{dR}{dt} \frac{\partial M}{\partial r} + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M,\end{aligned}\quad (31)$$

$$\begin{aligned}\frac{\partial v}{\partial t} = & \frac{D_v}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial v}{\partial r} \right) + \frac{r}{R} \frac{dR}{dt} \frac{\partial v}{\partial r} + \frac{d_y}{1 + k_y C} y \\ & - \delta_v v M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - \beta_v xv - d_v v,\end{aligned}\quad (32)$$

$$\begin{aligned}\frac{\partial P}{\partial t} = & \frac{D_P}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial P}{\partial r} \right) + \frac{r}{R} \frac{dR}{dt} \frac{\partial P}{\partial r} + \sigma_P \\ & + \lambda_P y \left(1 + \frac{\alpha_M M}{k_M + M} \right) - d_P P,\end{aligned}\quad (33)$$

$$\frac{\partial C}{\partial t} = \frac{D_C}{r^2} \frac{\partial}{\partial r} \left(\frac{r^2}{R^2} \frac{\partial C}{\partial r} \right) + \frac{r}{R} \frac{dR}{dt} \frac{\partial C}{\partial r} + \sigma_C + \lambda_C y - d_C C,\quad (34)$$

$$\begin{aligned}u = & \frac{R}{r^2} \int_0^r s^2 \left[\frac{\lambda_x}{1 + k_x C} x - d_n (1 - x - y - M) \right. \\ & \left. + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M \right] ds - \frac{\chi_P}{R} M \frac{\partial P}{\partial r} - \frac{\chi_C}{R} M \frac{\partial C}{\partial r}\end{aligned}\quad (35)$$

$$\frac{dR}{dt} = u(1, t).\quad (36)$$

З цим перетворенням рухома межева умова (рівняння (18)) стає рівнянням (36). Межові умови у центрі пухлини залишаються незмінними, так що

$$\frac{\partial x}{\partial r} = \frac{\partial y}{\partial r} = \frac{\partial n}{\partial r} = \frac{\partial M}{\partial r} = \frac{\partial v}{\partial r} = \frac{\partial P}{\partial r} = \frac{\partial C}{\partial r} = 0, \quad r = 0,\quad (37)$$

тоді як межові умови на межі пухлини, рівняння (21), (22), стають:

$$\begin{aligned}\frac{\partial x}{\partial r} = \frac{\partial y}{\partial r} = \frac{\partial n}{\partial r} = \frac{\partial M}{\partial r} = \frac{\partial v}{\partial r} = \frac{\partial P}{\partial r} = \frac{\partial C}{\partial r} = 0, \quad r = 1, \\ \frac{1}{R} \frac{\partial M}{\partial r} + \tilde{\alpha} \frac{P}{k_P + P} (M - \tilde{M}) = 0, \quad r = 1.\end{aligned}\quad (38)$$

Числова схема

Перепишемо безрозмірну модель (29)--(36) як

$$\frac{\partial z}{\partial t} + A_z \frac{\partial z}{\partial r} = \frac{D_z}{R^2} \frac{\partial^2 z}{\partial r^2} + F_z, \quad z = x, y, v, M, P, C. \quad (39)$$

$$\frac{1}{Rr^2} \frac{\partial(r^2 u)}{\partial r} = F, \quad (40)$$

$$\frac{dR}{dt} = u(1, t), \quad (41)$$

де $D_z = D$, $z = x, y, M$.

Коефіцієнти адвекції задані як

$$A_z = \frac{u}{R} - \frac{r}{R} \frac{dR}{dt} - \frac{2D_z}{R^2 r}, \quad z = x, y, M, \quad (42)$$

$$A_z = -\frac{r}{R} \frac{dR}{dt} - \frac{2D_z}{R^2 r}, \quad z = v, P, C. \quad (43)$$

$$F = \frac{\lambda_x}{1+k_x C} x - d_n(1-x-y-M) + \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M - \frac{1}{r^2} \frac{\partial}{\partial r} \left(\frac{\chi_P}{R^2} r^2 M \frac{\partial P}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(\frac{\chi_C}{R^2} r^2 M \frac{\partial C}{\partial r} \right), \quad (44)$$

$$F_x = \frac{\lambda_x}{1+k_x C} x - \beta_x x v - x F, \quad (45)$$

$$F_y = \beta_x x v - \delta_y y M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - \frac{d_y}{1+k_y C} y - y F, \quad (46)$$

$$F_M = \lambda_M \left(1 + \frac{\alpha_{MC} C}{k_{MC} + C} \right) - d_M M - M F - \frac{1}{r^2} \frac{\partial}{\partial r} \left(\frac{\chi_P}{R^2} r^2 M \frac{\partial P}{\partial r} \right) - \frac{1}{r^2} \frac{\partial}{\partial r} \left(\frac{\chi_C}{R^2} r^2 M \frac{\partial C}{\partial r} \right), \quad (47)$$

$$F_v = \frac{d_y}{1+k_y C} y - \beta_x x v - \delta_v v M \left(1 + \frac{\alpha_C C}{k_C + C} \right) - d_v v, \quad (48)$$

$$F_P = \sigma_C + \lambda_P y \left(1 + \frac{\alpha_M M}{k_M + M} \right) - d_P P, \quad (49)$$

$$F_C = \sigma_C + \lambda_C y - d_C C. \quad (50)$$

Розділ 10. Моделювання інвазії пухлин

10.1 Компоненти системи uPA та їх функції

Протеолітична деградація та ремоделювання позаклітинного матриксу є важливими для інвазії ракових клітин. Це дозволяє раковим клітинам проліферувати та мігрувати через навколишні тканини. У цьому контексті одним із перших кроків вторгнення є виробництво і секреція протеолітичних ферментів, тобто активатора урокінази плазміногену (uPA) і матриксних металопротеїназ (ММР) раковими клітинами. Ці ферменти взаємодіють з динамікою макромолекул ЕСМ і прокладають шлях до інвазії раку. Зокрема, ферментативна система uPA в основному складається з рецептора урокінази (uPAR), активатору плазміногену урокінази (uPA), матриксноподібного білку вітрoneктин (VN), інгібітору активатора плазміногену типу 1 (PAI-1) і ферменту плазміну, що розкладає. Тут ми перелічуємо ключові види діяльності основних компонентів системи uPA та роль кожного з них в інвазії раку.

10.1.1 Активатор плазміногену урокінази (uPA)

uPA є позаклітинна серинова протеаза, що виробляється клітинами. Два основних функціональних областей молекули uPA є області протеази та фактора росту. Протеазна частина активує плазміноген і перетворює його на плазмін, який здатний перетравлювати білки базальної мембрани та позаклітинного матриксу. Область фактору росту не має протеазної активності, але може зв'язувати специфічний високоафінний рецептор клітинної поверхні, uPAR. Нарешті, uPA має зимогенну форму, pro-uPA, яка може бути активована плазміном і зв'язується з uPAR.

10.1.2 Рецептор активатора плазміногену урокінази (uPAR)

uPAR є високоафінним клітинним поверхневим рецептором uPA (і його зимогенної форми pro-uPA), який через процес зв'язування локалізує uPA та pro-uPA на поверхні клітини. Важливо, що uPAR містить інше місце зв'язування для компонента ЕСМ під назвою вітрoneктин (VN), і оскільки сайти зв'язування VN та uPA є різними, uPAR може одночасно зв'язувати обидва ліганди, забезпечуючи узгоджену регуляцію протеолізу, клітинної адгезії та сигналізації.

10.1.3 Вітрoneктин (VN)

VN - це багатофункціональний глікопротеїн, який міститься у сироватці крові та ЕСМ і сприяє адгезії та поширенню клітин. Відповідно до назви, вітрoneктин міцно зв'язується зі скляними поверхнями (vitro = скло), і він має сайти зв'язування для кількох лігандів, включаючи гепарин, *урокіназний рецептор активатора плазміногену (uPAR)*, інгібітор активатора *плазміноген тип-1 (PAI-1)* та інтегрини, такі як $\alpha_v\beta_3$.

10.1.4 Інгібітор активатора плазміногену урокінази-1 (PAI-1)

Один з інгібіторів урокінази, активатор плазміногену PAI-1 належить до серпіну (інгібітори серинового білка), і вважається, що воно є найпоширенішим, швидкодіючим інгібітором uPA *in vivo*. Він може специфічно зв'язуватися з розчинним і пов'язаним з мембраною uPA для пригнічення активації плазміногену. Коли PAI-1 зв'язується з комплексом uPA/uPAR, він запускає інтерналізацію комплексу uPA/uPAR/PAI-1 за допомогою рецепторів ендцитоз, що означає, що комплекс буде дисоційовано, а PAI-1 і uPA будуть перетравлюватися, але рецептор повертається на поверхню клітини. Цей процес допомагає при очищенні PAI-1 від поверхні клітини.

10.1.5 Плазмін

Плазмін є широко поширеним ферментом, який розщеплює багато білків позаклітинного матриксу, таких як фібронектин, ламінін, вітронектин і тромбоспондин. Крім того, плазмін також може активувати багато матричних металопротеїназ (ММР), ще більше посилюючи деградацію позаклітинного матриксу.

10.2 Математичне моделювання

Розглянемо три основні компоненти багатомасштабної моделі uPA, що пропонується для інвазії раку, а саме макроскопічну динаміку, мікроскопічну динаміку та регуляцію переміщення меж пухлини.

10.2.1 Макроскопічна динаміка

Густину ракових клітин позначимо через c , густину позаклітинного матриксу через v (без розрізнення ЕСМ та його компоненту VN), концентрацію активатора плазміногену урокінази (uPA) -- u , концентрацію інгібітору активатора плазміногену (PAI-1) -- p і концентрацію плазміну -- m .

Розглядається концентрація uPAR пропорційно густині ракових клітин. Ще одне важливе припущення полягає у тому, що запас плазміногену у цій моделі необмежений. Нарешті, макроскопічна модель отримана із урахуванням біологічних міркувань, описаних у попередньому розділі у поєднанні з такими презумпціями:

Динаміка ракової клітини.

Передбачається, що міграція ракових клітин в основному регулюється дифузією, хемотаксисом через uPA та PAI-1 і гаптотаксис через VN та інші компоненти ЕСМ. Крім того, використовується закон логістичного зростання для моделювання проліферації ракових клітин. Отже, рівняння для густини ракових клітин, яка тут розглядається, є:

$$\frac{\partial c}{\partial t} = D_c \Delta c - \nabla \cdot [\chi_u c \nabla u + \chi_p c \nabla p + \chi_v c \nabla v] + \mu_1 c \left(1 - \frac{c}{c_0} \right), \quad (1)$$

де D_c -- коефіцієнт дифузії ракових клітин, χ_u і χ_p -- коефіцієнти хемотаксису відносно uPA та PAI-1, відповідно, χ_v -- швидкість VN-опосередкованого гаптотаксису, μ_1 -- швидкість проліферації ракових клітин, а c_0 -- максимальна пропускна здатність для ракових клітин.

Динаміка ЕСМ/VN.

Оскільки не передбачається переміщення ЕСМ, ми виключаємо будь-які умови

міграції. Тоді як VN (який є важливим компонентом ECM) погіршується при контакті з ферментами m , зв'язування PAI-1 з uPA пригнічує активацію плазміногену, що призводить до захисту VN та інших молекулярних складових ECM та опосередковано сприяє їх виробництву. Одночасно відбувається зв'язування PAI-1 з VN у меншому зв'язуванні з рецепторами клітинної поверхні, такими як uPAR, і, отже, через регуляцію шляхів передачі сигналу, асоційованих із клітинною матрицею, це пригнічує виробництво VN. Отже, припускаючи логістичну реконструкцію ECM, маємо рівняння для ECM:

$$\frac{\partial v}{\partial t} = -\delta v m + \phi_{21} u p - \phi_{22} v p + \mu_2 v \left(1 - \frac{v}{v_0}\right), \quad (2)$$

де δ – швидкість деградації ECM плазміном, ϕ_{21} – швидкість зв'язування PAI-1 до uPA, ϕ_{22} – швидкість зв'язування PAI-1 з VN, а μ_2 – швидкість ремоделювання матриці.

Динаміка uPA.

Математичне моделювання динаміки концентрації uPA враховується для таких аспектів. Коли вона виробляється раковими клітинами і видалається із системи через її зв'язування з PAI-1 і uPAR, за одиницю часу uPA здійснює локальну дифузію. Тому математично це можна формалізувати так:

$$\frac{\partial u}{\partial t} = D_u \Delta u - \phi_{31} u p - \phi_{33} c u + \alpha_{31} c, \quad (3)$$

де D_u – коефіцієнт дифузії, ϕ_{31} і ϕ_{33} – швидкості зв'язування uPA/PAI-1 і uPA/uPAR відповідно, а α_{31} – швидкість виробництва uPA раковими клітинами.

Динаміка PAI-1.

Аналогічно рівняння для PAI-1 просто включає член дифузії з коефіцієнтом D_p , видалення, спричинене зв'язуванням з uPA та VN зі швидкістю зв'язування ϕ_{41} і ϕ_{42} відповідно, і виробництво як результат утворення плазміну зі швидкістю α_{41} .

$$\frac{\partial p}{\partial t} = D_p \Delta p - \phi_{41} u p - \phi_{42} p v + \alpha_{41} m \quad (4)$$

Динаміка плазміну.

Еволюція концентрації плазміну моделюється наступним чином. Припускаючи, що за одиницю часу це здійснює локальну дифузію, ми вважаємо, що зв'язування uPA з uPAR забезпечує можливість перичелюлярної протеолітичної активності через активацію плазміногену, що призводить до утворення плазміну. Більш того, зв'язування PAI-1 з VN опосередковано посилює зв'язування uPA з uPAR, таким чином вносячи додатковий внесок в утворення плазміну. Отже, ці припущення дають нам такий закон еволюції:

$$\frac{\partial m}{\partial t} = D_m \Delta m + \phi_{52} p v + \phi_{53} c u - \phi_{54} m \quad (5)$$

де D_m – коефіцієнт дифузії, ϕ_{52} і ϕ_{53} – швидкості зв'язування PAI-1/VN і uPAR/uPA відповідно, а ϕ_{54} – швидкість розпаду плазміну.

10.2.2 Мікроскопічна динаміка

Звертаючись тепер до налаштування мікромасштабу, ми виведемо наступну систему з трьох пов'язаних PDE для опису мікродинаміки активації плазміногену, що відбувається в клітинному масштабі ε -околиці $\mathcal{P}$ інвазивного краю пухлини (що

запроваджено в Додатку 1).

Припускаючи, що PAI-1 і uPAR рівномірно експресуються на поверхні клітини різні типи клітин у пухлині, ця система буде фіксувати мікродинаміку uPA, PAI-1 і плазміну, враховуючи такі біологічні міркування. З одного боку, передбачається урокіназний активатор плазміногену (uPA). для зв'язування з поверхневим рецептором раку uPAR для активації плазміногену, що призводить до деградації перичелюлярного ЕСМ через серію протеолітичних активностей. З іншого боку, пов'язані з мембраною ММР (такі як MT1-ММР) секретуються зсередини популяції пухлинних клітин розподілентх на зовнішньому проліферуючому обідку вздовж усієї периферії пухлини. Отже, на основі цих міркувань ми пропонуємо пов'язаний керівний закон для переднього краю мікродинаміки, яка детально описується так:

Мікродинаміка uPA

У кожній мікрообласті εY динаміка молекулярної популяції uPA регулюється шляхом дифузійного процесу, джерелом якого є макродинаміка пухлинної клітини. У кожній точці $y \in \varepsilon Y$ джерело uPA виникає як сукупний внесок пухлинних клітин, розподілених у певній сусідній області всередині зовнішньої частини пухлини проліферуючого обідку. Тому це джерело позначається $f_{\varepsilon Y(\cdot)} \times \varepsilon Y \rightarrow \mathbb{R}_+$ і визначається як:

$$f_1^{\varepsilon Y}(y, \tau) \begin{cases} \frac{1}{\lambda[B(y, \gamma) \cap \Omega(t_0)]} \int_{B(y, \gamma) \cap \Omega(t_0)} c(x, t_0 + \tau) dx, & y \in \varepsilon Y \cap \Omega(t_0), \\ 0, & \text{поза раку,} \end{cases} \quad (6)$$

де $\lambda(\cdot)$ — стандартна міра Лебега, а γ — максимальна товщина зовнішнього проліферуючого обідку. Таким чином, за одиницю часу, за наявності джерела (6) uPA локально дифундує і зв'язується як з PAI-1, так і з uPAR, тому його мікродинаміку можна формально записати так:

$$\frac{\partial u}{\partial \tau} = D_u \Delta u - \phi_{31} p u + (\alpha_{31} - \phi_{33} u) f_1^{\varepsilon Y}(y, \tau). \quad (7)$$

Мікродинаміка PAI-1

Рівняння для PAI-1 враховує дифузійний рух, виробництво за рахунок активації плазміну та втрати внаслідок зв'язування з uPA та VN. Зокрема, прив'язка між PAI-1 і VN є сукупним ефектом розподілу ЕСМ в межах εY . Тому діючи аналогічно до випадку вихідного члена в ((143)), ми визначимо

$$f_2^{\varepsilon Y}(y, \tau) = \frac{1}{\lambda[B(y, 2\varepsilon)]} \int_{B(y, 2\varepsilon)} v(x, t_0 + \tau) dx, \quad y \in \varepsilon Y. \quad (8)$$

які надають можливість написати керуючий закон для мікродинаміки PAI-1, а саме:

$$\frac{\partial p}{\partial \tau} = D_p \Delta p - \phi_{41} p u + \alpha_{41} m - \phi_{42} p f_2^{\varepsilon Y}(y, \tau). \quad (9)$$

Мікродинаміка плазміну

Для просторово-часової еволюції плазміну передбачається, що за одиницю часу він здійснює локальну дифузію за наявності наступного джерела та обставин розпаду. Враховуючи, що зв'язування uPA з uPAR необхідне для забезпечення поверхні клітини з потенційною протеолітичною активністю, з одного боку, джерело плазміну на внесок

зв'язування uPA/uPAR. З іншого боку, оскільки PAI-1 разом конкурує з uPAR за зв'язування з VN, зв'язування PAI-1 з VN дає більше можливості uPAR зв'язуватися з uPA і опосередковано призводить до більшого утворення плазміну. Нарешті, плазмін може бути дезактивованій або шляхом деградації, або під дією інгібітору плазміну α_2 -антиплазміну. Отже, рівняння, що описує ці біологічні взаємодії, це:

$$\frac{\partial m}{\partial \tau} = D_m \Delta m - \phi_{5,3} m + \phi_{52} p f_2^{\varepsilon Y}(y, \tau) + \phi_{53} u f_1^{\varepsilon Y}(y, \tau). \quad (10)$$

10.2.3 Макроскопічне переміщення межі пухлини

Відповідно до багатомасштабного підходу, описаного в Додатку 1, набір точок $\{x_{\varepsilon Y}^*\}$ на межі пухлини у поточний момент часу рухається у бік набору нових просторових позицій $\{x_{\varepsilon Y}^*\}$ для формування нової межі на наступному багатомасштабному етапі за умови, що локальна перехідна ймовірність q^* узгоджується з обставинами у навколишнє перитуморальне мікрооточення. Коли сила вторгнення вище тканинного порогу, точка $x_{\varepsilon Y}^*$ переміститься у нове положення $x_{\varepsilon Y}^*$ за напрямком і величина зміщення, яка представляє рух усіх точок від частини інвазивного краю, захопленого мікрообластю εY .

На будь-якій мікрообласті εY за умови, що була вироблена достатня кількість плазміну через вторгнення краю, це картина фронту просування просторового розподілу плазміну, що характеризує деградацію ЕСМ. Отже, напрямок руху і величина переміщення частини вторгнення краю пухлини, захопленого поточною мікрообластю εY , визначатиметься просторовою схемою розподілу наступаючого фронту плазміну $m(\cdot, \tau_f)$ у навколопухлинній області. Ці характеристики руху отримані з урахуванням внеску усіх піків (з барицентром у просторових точках y_l) на передній частині плазміну, що просувається, вище середнього значення всієї маси плазміну, що виробляється на $\varepsilon Y \setminus \Omega(t_0)$ і міститься на найдальшій евклідовій відстані від $\{x_{\varepsilon Y}^*\}$. Отже, за цих умов напрямок руху $\eta_{\varepsilon Y}$ та величина зміщення $\xi_{\varepsilon Y}$ (Додаток 1) у (35)–(36), мають такі вирази:

$$\eta_{\varepsilon Y} = x_{\varepsilon Y}^* + \nu \sum_{l \in T_\delta} \int_{D_l} m(y, \tau_f) dy (y_l - x_{\varepsilon Y}^*), \quad \nu \in [0, \infty),$$

$$\xi_{\varepsilon Y} = \sum_{l \in T_\delta} \frac{\int_{D_l} m(y, \tau_f) dy}{\sum_{l \in T_\delta} \int_{D_l} m(y, \tau_f) dy} |x_{\varepsilon Y}^* y_l|.$$

Нарешті, ймовірність переходу q^* , є кількісним визначенням суми плазміну в $\varepsilon Y \setminus \Omega(t_0)$ відносно загальної кількості концентрації плазміну в εY і характеризує силу вторгнення. Отже, точка $x_{\varepsilon Y}^*$ здійснюватиме переміщення в нове просторове положення $x_{\varepsilon Y}^*$ тоді і тільки тоді, коли $q(x_{\varepsilon Y}^*) = q^*(\varepsilon Y)$ перевищує певний тканинний локальний поріг $\omega(\beta, \varepsilon Y) \in (0, 1)$, асоційований з мікрообластю εY за заданого стану сприятливих умов β .

Табл.10.1. Параметри в $\mathcal{P}$

Параметр	Значення	Опис
D_c	$4,3 \cdot 10^{-3}$	Дифузія пухлинних клітин
χ_u	0,0305	Хемотаксис до uPA
χ_p	0,0375	Хемотаксис до PAI-1
χ_v	0,0285	Гаптотаксис до ЕСМ (вітронектину)
μ_1	0,25	Проліферація ракових клітин
δ	1,5	Деградація ЕСМ
ψ_{21}	0,75	Зв'язування uPA і PAI-1
ψ_{22}	0,55	Зв'язування PAI-1 та VN
μ_2	0,01	Проліферація ЕСМ
D_u	$2,5 \cdot 10^{-3}$	Дифузія uPA
ψ_{31}	0,75	Зв'язування uPA і PAI-1
ψ_{33}	0,3	Зв'язування PAI-1 та uPAR
α_{31}	0,215	Проліферація uPA
D_p	$3,5 \cdot 10^{-3}$	Дифузія PAI-1
ψ_{41}	0,75	Зв'язування uPA і PAI-1
ψ_{42}	0,55	Зв'язування PAI-1 та VN
α_{41}	0,5	Проліферація PAI-1
D_m	$4,91 \cdot 10^{-3}$	Дифузія плазміну
ψ_{52}	0,11	Збільшення швидкості
		за рахунок зв'язування PAI-1 і VN
ψ_{53}	0,75	Збільшення швидкості
		за рахунок зв'язування uPA та uPAR
ψ_{54}	0,5	Деградація плазміну

10.3 Результати моделювання

Запропонована багатомасштабна модель інвазії раку була розв'язана чисельно у прямокутній області $Y = [0, 4] \times [0, 4]$. Для всіх наших моделювань ми дискретизуємо весь куб Y рівномірно, використовуючи розмір просторової сітки: $\Delta x = \Delta y = \varepsilon/2 = 0,03125$. У цьому контексті початкові умови для макродинаміки задаються наступним чином. Популяція ракових клітин на спочатку розглянутій ділянці пухлини $\Omega(0) = B[(2, 2), 0.5]$ припускається, що задано перекладеним гауссовим виразом з центром $(2, 2)$ і пом'якшеним до 0 за радіусом 0,5, а саме

$$c(x, 0) = \frac{1}{2} \left[\exp \left(-\frac{\|x - (2, 2)\|_2^2}{\sqrt{\Delta x \Delta y}} \right) - \exp(-28,125) \right] (\chi_{B[(2,2),0.5-y]} * \Psi_\gamma), \quad x \in Y. \quad (11)$$

де Ψ_γ – пом'якшувач, з $\gamma \ll \Delta x/3$. Далі, початкові макроскопічні умови для ферментативних компонентів, що надходять в систему uPA, розглядаються як:

$$u(x, 0) = 1 - \frac{1}{2} c(x, 0), \quad p(x, 0) = \frac{1}{2} c(x, 0), \quad m(x, 0) = \frac{1}{20} c(x, 0). \quad (12)$$

Далі розглядаються такі однорідна та неоднорідна початкові умови для ЕСМ, а саме: однорідний випадок:

$$v(x, 0) = 1 - c(x, 0), \quad x \in Y. \quad (13)$$

неоднорідний випадок:

$$v(x, 0) = \frac{1}{2} [1 + 0.3 \sin(4\pi \|x\|_2) + \sin(4\pi m \|(4, 0) - x\|_2)] \quad (14)$$

На рис. 10.1 показано початкові умови (11) і (14) розподілу ЕСМ і ракових клітин як для однорідних, так і для гетерогенних випадків.

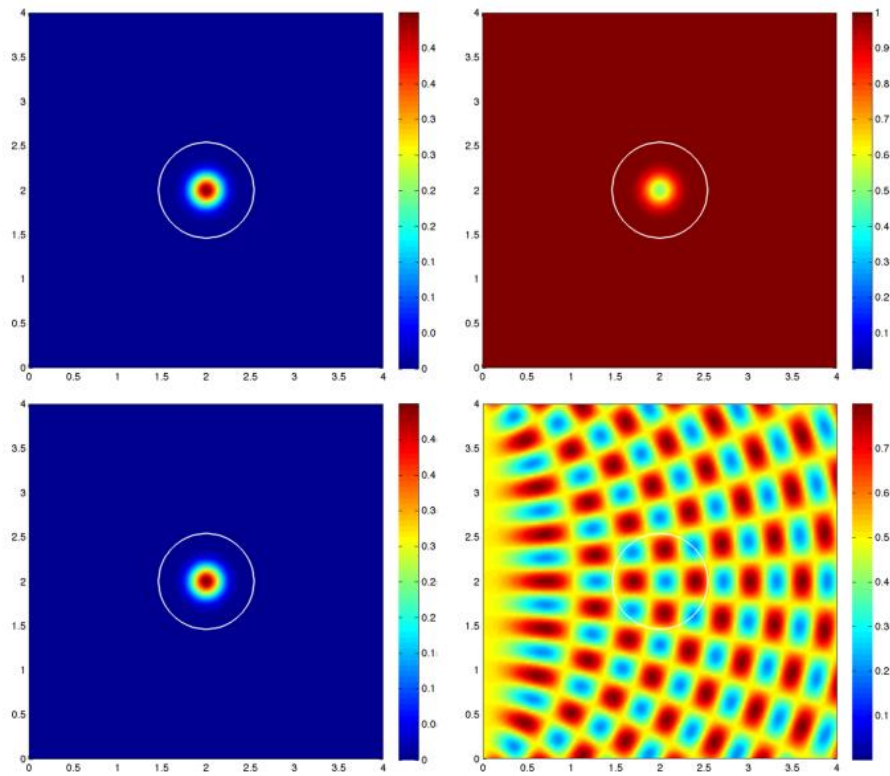


Рис.10.1. Початкові умови розподілу пухлинних клітин (ліворуч та ЕСМ (праворуч) та інвазивна межа

Нарешті, за відсутності медичних даних, для тканинного порогу $\omega(\beta, \varepsilon Y)$, що контролює чи буде точка на межі здійснювати рух відповідно до напрямку і величини зміщення, визначені в Додатку 1, ми приймаємо наступну функціональну форму:

$$\omega(\beta, \varepsilon Y) = \begin{cases} \sin \left[\frac{\pi}{2} \left(1 - \frac{1}{\beta} G(\beta, \varepsilon Y) \right) \right], & G(\beta, \varepsilon Y) \leq \beta, \\ \sin \left[\frac{\pi}{2(1-\beta)} (G(\beta, \varepsilon Y) - \beta) \right], & G(\beta, \varepsilon Y) > \beta, \end{cases} \quad (15)$$

$$G(\beta, \varepsilon Y) = \frac{v_{\omega(t_0)}(x_{\varepsilon Y}^*, t_0 + \Delta t)}{\sup_{\xi \in \partial\Omega(t_0)} v_{\Omega(t_0)}(\xi, t_0 + \Delta t)},$$

де $\beta \in (0,1)$ – параметр, який контролює певний «оптимальний рівень» деградації ЕСМ і розглядають це як показник найсприятливіших умов вторгнення на рівні пухлини і тканинного мікрооточення. Ця функціональна форма правил ω будь-якого вторгнення, якщо повне знищення або дуже поверхнева деградація навколишній ЕСМ виконується за допомогою MDE. Це пов'язано з тим, що поки для інвазії ракові клітини потребують значного рівня деградації ЕСМ, загалом руйнування навколишньої структури ЕСМ заважає їм просуватися далі у тканину, оскільки їм потрібні деякі з компонентів ЕСМ, щоб прилипнути до них, щоб мігрувати.

У подальшому моделюванні для макроскопічної частини моделі, якщо не вказано інше, ми будемо використовувати наступний базовий набір значень параметрів $\mathcal{P}$:

$$D_n = 4,3 \cdot 10^{-3}, \chi_u = 0,0305, \chi_p = 0,0375, \chi_v = 0,0285,$$

$$\mu_1 = 0,25, \mu_2 = 0,15, \delta = 1,5,$$

$$\phi_{21} = 0,75, \phi_{22} = 0,55, \phi_{31} = 0,75, \phi_{33} = 0,3, \phi_{41} = 0,75, \phi_{42} = 0,55, \phi_{52} = 0,11,$$

$$\phi_{53} = 0,75; \phi_{54} = 0,5, \alpha_{31} = 0,5, \alpha_{41} = 0,5,$$

$$D_u = 2,5 \cdot 10^{-3}, D_p = 3,5 \cdot 10^{-3}, D_m = 4,91 \cdot 10^{-3}.$$

які детально описані в табл. 9.1 додатку 3. Однак, щоб дослідити різні моделі росту раку, як описано у наступному моделюванні, ми виконаємо числові тести також для дещо змінених значення коефіцієнта дифузії ракових клітин D_c , швидкість проліферації ЕСМ або швидкість деградації ЕСМ ?. Крім того, ми дослідимо діапазон значень параметра β , щоб підкреслити кореляцію між змінами умов мікрооточення тканини і результуючі схеми інвазії раку.

Щоб проаналізувати вплив кожної змінної (а саме початковий стан ЕСМ, коефіцієнт дифузії ракових клітин, швидкість проліферації та деградації ЕСМ, а також поріг коефіцієнт), ми розділимо результати на чотири групи.

10.3.1 Початковий стан ЕСМ

Щоб дослідити який вплив мають різні початкові умови ЕСМ на загальній динаміці моделі ми застосовуємо однакову порогову функцію для обох однорідного і неоднорідного сценаріїв ЕСМ.

Гетерогенність ЕСМ призводить до природної взаємодії між пухлиною та тканиною складніше, ніж у випадку однорідного ЕСМ. Прямий наслідок цього полягає у тому, що неоднорідність ЕСМ викликає відповідну внутрішню мінливість у тканинні пороги, які досліджують стан навколопухлинної тканини, що викликає взаємодію раку під час вторгнення. Це призводить до вищого ступеня складності у просторовій структурі областей з найсприятливішими умовами для пухлини просуватися далі у навколишній області, у межах напрямків і переміщення величини, визначеної мікромасштабною динамікою. Зрештою, це породжує вищий рівень аплікатури та інфільтративних візерунків у гетерогенному випадку на противагу до однорідного випадку. Це підтверджується нашими результатами, представленими на рис. 9.3, отриманих для гетерогенних ЕСМ, які виявляють більш «пальцеве» та інфільтративне поширення пухлини порівняно з результатами, отриманими у випадку гомогенного ЕСМ рис. 9.2. Цей тип візерунків

пальців часто спостерігається при медичній візуалізації даних

10.3.2 Коефіцієнт дифузії ракових клітин D_c

Умови хемотаксису в раковій клітині (1) є основними причинами виникнення неоднорідних візерунків всередині пухлини. Отже, якщо коефіцієнт дифузії ракових клітин (D_c) збільшити на один порядок більше, ніж коефіцієнти хемотаксису (χ_u і χ_p), що стає домінуючим механізмом руху клітини, тоді гетерогенної динаміки всередині пухлини не буде, як показано на рис. 9.4, 9.5, коефіцієнти хемотаксису на порядок більші за коефіцієнт дифузії, і, як наслідок, ми отримуємо неоднорідну картину формування раку клітин, що призводить до динамічнішої деформації меж пухлини.

10.3.3 Швидкість розповсюдження ЕСМ μ_2 і швидкість деградації

δ

З усіх представлених досі результатів моделювання ми робимо висновок, що деградація ЕСМ полегшує інвазію раку. Однак, як показано також у кількох міждисциплінарних біологічних дослідженнях, процес інвазії зупиниться там, де ЕСМ сильно погіршиться. Це фіксує біологічний сценарій у 2D, коли також спостерігається адгезія клітинної матриці низький, не утворюються вогнищеві спайки або стресові волокна, клітини не рухаються.

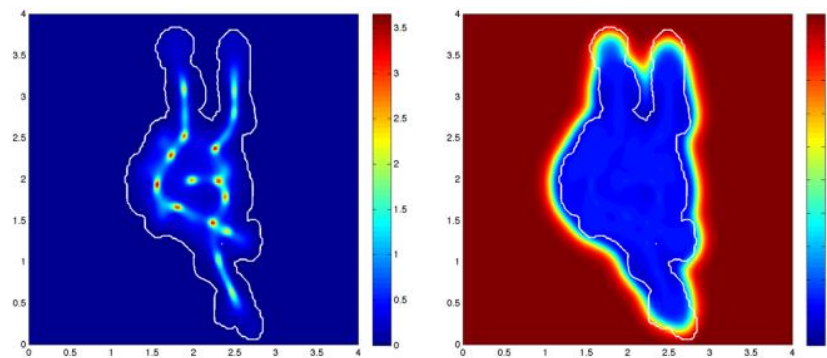


Рис. 10.2 Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на різних макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,775$, $\mu_2 = 0,01$, $\delta = 1,5$

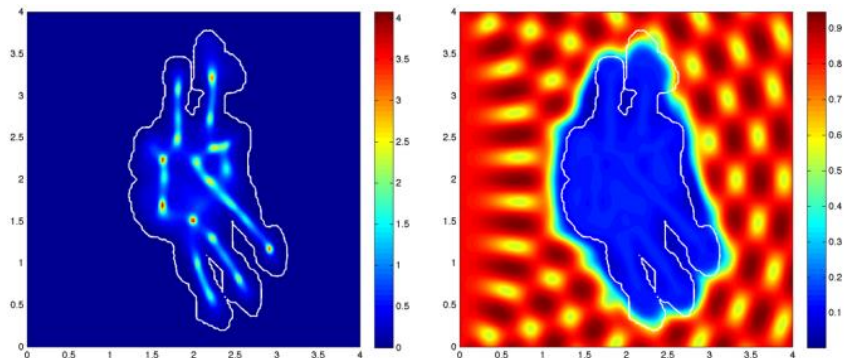


Рис.10.3 Результати моделювання, що показують розподіл ракових клітин (лівий

стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. 10.1, ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,775$, $\mu_2 = 0,01$, $\delta = 1,5$

Щоб дослідити вплив проліферації та деградації ЕСМ на процес інвазії, порівняємо дві групи параметрів: (1) $\mu_2 = 0,01$, $\delta = 1,5$ (швидкість проліферації з відносно високою швидкістю деградації, рис.6); (2): $\mu_2 = 0$, $\delta = 0,75$ (відсутність проліферації з відносно невеликою швидкістю деградації, рис. 9.7). З цих двох груп зображень, ми спостерігаємо, що коли проліферація присутня, і швидкість деградації є порівняно великі, деформації кордону не такі динамічні, як при термін проліферації відсутній з відносно невеликою швидкістю деградації. Причина може бути в тому, що термін проліферації зменшить ступінь неоднорідності розподілу ЕСМ, що призводить до меншого поширення популяції ракових клітин.

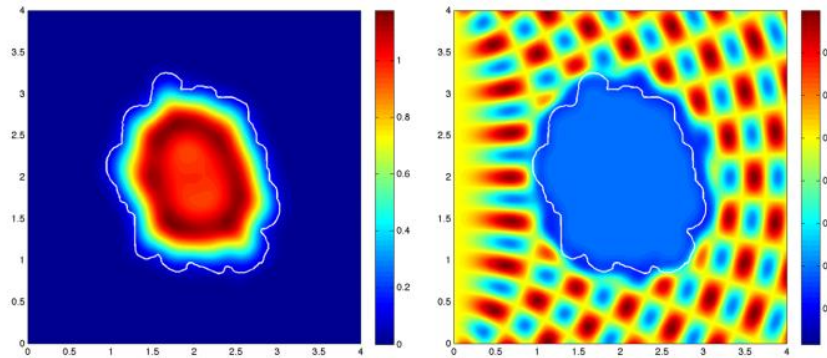


Рис. 10.4 Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,775$, $\mu_2 = 0,0$, $\delta = 0,75$

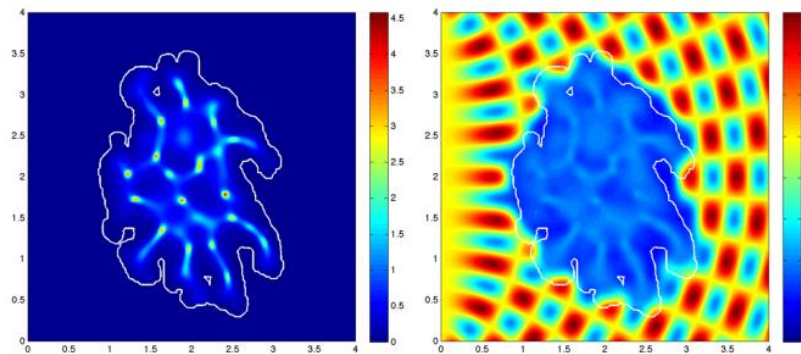


Рис. 10.5. Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,775$, $\mu_2 = 0,0$, $\delta = 0,75$

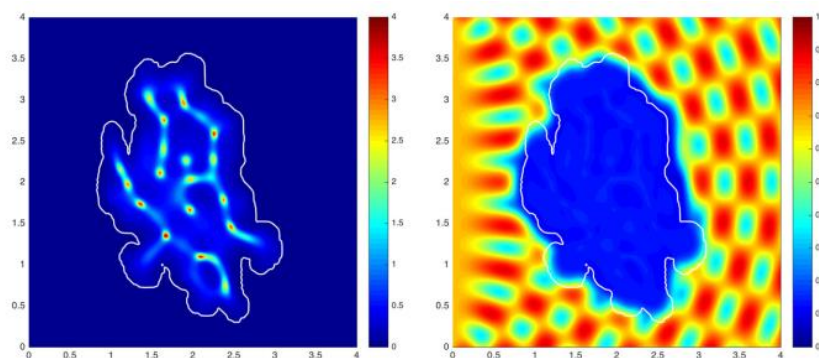


Рис.10.6. Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,7625$, $\mu_2 = 0,005$, $\delta = 1,5$

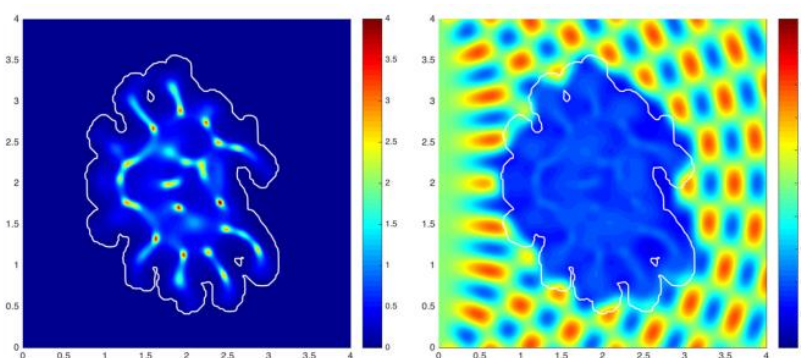


Рис. 10.7. Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,7625$, $\mu_2 = 0,0$, $\delta = 0,75$

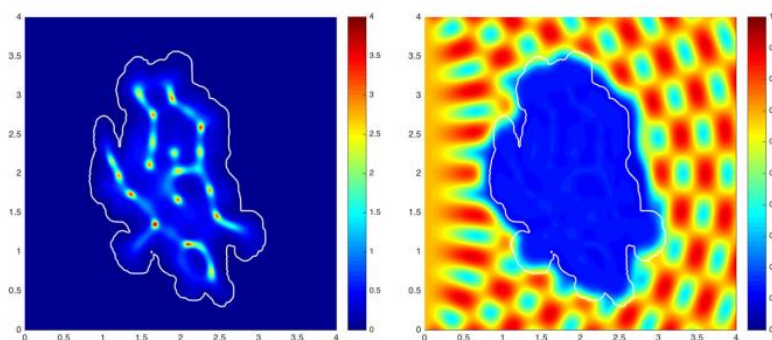


Рис. 10.8. Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. 10.1, ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,7625$, $\mu_2 = 0,005$, $\delta = 1,5$

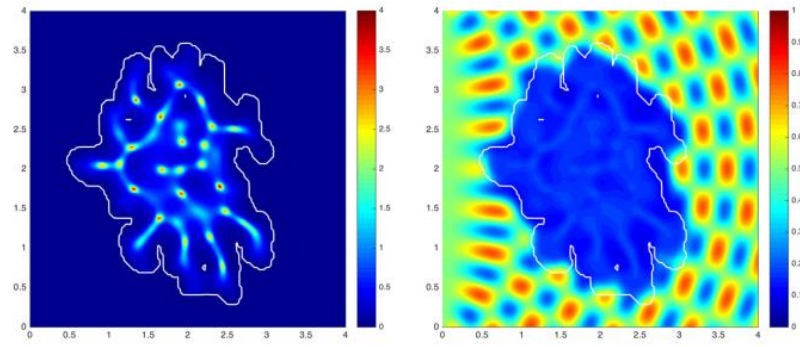


Рис. 10.9. Результати моделювання, що показують розподіл ракових клітин (лівий стовпець) і ЕСМ (правий стовпець) і інвазивна межа пухлини (біла лінія) на макро-мікро стадії 60. Початок з однорідних початкових умов, показаних на рис. (156), ці результати були отримані для $D_c = 4,3 \cdot 10^{-3}$, $\beta = 0,7625$, $\mu_2 = 0,0$, $\delta = 1$

Розділ 11 Комбінована терапія меланоми

11.1 Вступ

PD-1 є переважно імунноінгібіторним рецептором експресується на активованих Т-клітинах. Його ліганд PD-L1 посилено регулюється на тих самих активованих Т-клітинах, і також експресується мієлоїдними супресорними клітинами (MDSCs) [p122] і в деяких ракових клітинах людини, в т.ч. меланомі, раку легень, раку товстої кишки та лейкозії. Відомо, що комплекс PD-1-PD-L1 пригнічує функцію Т-клітин [p122]. Імунні контрольні точки є регуляторними шляхами імунної системи, яка пригнічує її активну реакцію проти конкретних цілі. При онкологічних захворюваннях комплекс PD-1-PD-L1 функціонує як імунна контрольна точка для протипухлинної боротьби Т-клітини. За останні роки було досягнуто значного прогресу у розробці інгібіторів контрольних точок, насамперед анти-PD-1 і анти-PD-L1.

Ген B-raf мутований у 66% злоякісних меланом людини, а його білковий продукт, BRAF-кіназа, є ключовою частиною шляху RAS-RAF-MEK-ERK (MAPK) проліферації ракових клітин [F1]. BRAF терапія викликає значну відповідь у більшості пацієнтів, але відповідь короткочасна (близько 6 місяців) [F1]. Початкові клінічні дослідження свідчать, що одночасне інгібування BRAF з MEK зменшує набуту резистентність, викликану MAPK, що призводить до підвищення клінічної ефективності та зниження токсичності [F2]. Це забезпечує раціональне використання комбінованого інгібування BRAF/MEK замість інгібування BRAF окремо. Лікування меланоми інгібіторами імунних контрольних точок має нижчий рівень відповіді порівняно з лікуванням інгібіторами BRAF/MEK, але відповідь має тенденцію бути міцнішою, зберігається роками [F10].

Метааналіз рандомізованих клінічних досліджень показує, що порівняно з іншими методами лікування прогресуючої мутації BRAF меланома, комбіноване інгібування BRAF/MEK та PD-1 значно покращився загальний час виживання.

Фактори, викликані меланою, змінюють дозрівання і активацію тканинно-резидентних дендритних клітин, сприяючи таким чином імунному зникненню пухлини [F7]. У мутантній меланомі BRAF, інгібітор BRAF відновлює пошкоджену функцію дендритної клітини.

IL-12, отриманий з дендритних клітин, активує ефекторні Т-клітини (Th1 і $CD8^+$ Т-клітини). Th1 продукує IL-2, що додатково сприяє проліферації ефекторних Т-клітин. Т-клітини $CD4^+$ Т-клітини (Th1) можуть безпосередньо вбивати ракові клітини, наприклад, через FAS- або TRAIL-залежний шлях [F8], тоді як $CD8^+$ Т-клітина ефективніша у знищенні ракових клітин. Ракові клітини пригнічують функції ефекторних Т-клітин, виробляючи імуносупресорні цитокіни TGF- β , IL-6, CCL2 та IL-10 [F11]. IL-10 пригнічує активацію $CD8^+$ Т-клітин і Th1 [F11].

Одним із контрольних пунктів Т-клітин є білок PD-1 мембрани. Його ліганд PD-L1 експресується на активованому ефекторні Т-клітини, на MDSC та ракові клітини

Вищевказані взаємодії між раковими клітинами та імунних клітин узагальнено на рис. 10.1. Математичні модель базується на рис. 10.1, і містить інгібітори BRAF/MEK та PD-1.

11.2 Математична модель

Математична модель базується на показаній мережі на рис.10.1. Перелік змінних наведено в табл. **Ошибка! Источник ссылки не найден..** Оскільки CCL2 та IL-6 виробляються раковими клітинами і обидва передаються MDSC та Tregs у пухлинне середовище, у моделі розглянуто, для простоти, лише IL-6.

Ми припускаємо, що загальна густина клітин всередині пухлини залишається постійною у просторі і часі:

$$C + D + T_1 + T_8 + T_r + M = \text{const.} \quad (1)$$

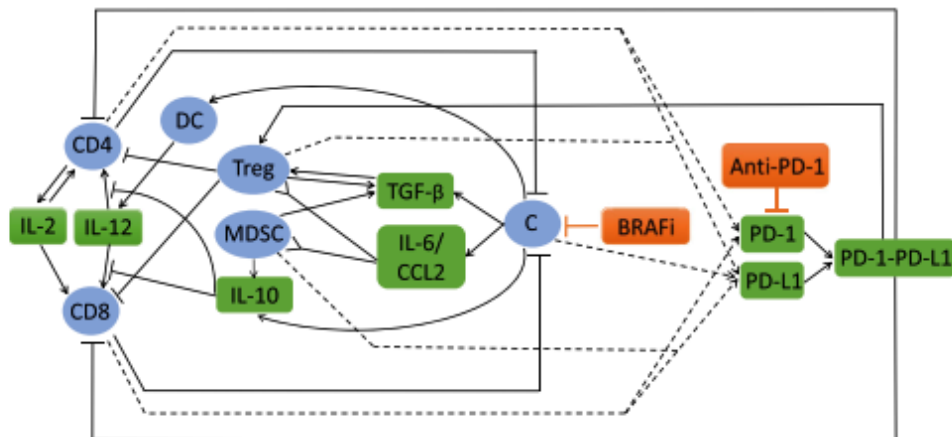


Рис. 11.1. Взаємодія імунних клітин із раковими

Гострі стрілки вказують на поширення/активацію, заблоковані стрілки вказують на вбивство/блокування, перевернуті гострі стрілки вказують на рекрутинг/ хіміоатракцію, а пунктирні лінії вказують на білки на Т-клітинах, MDSC та ракових клітинах.

Табл. 11.1 Параметри моделі (g/cm^3)

Параметр	Опис
H	HMGB-1 концентрація
N_c	густина некротичних ракових клітин
D	густина ДК
T_1	густина активованих CD4+ Т-клітин
T_8	густина активованих CD8+ Т-клітин
T_r	густина активованих Treg-клітин
M	густина активованих МДСК
C	густина ракових клітин
I_{12}	Концентрація IL-12
I_2	Концентрація IL-2
T_β	Концентрація $TGF - \beta$
I_6	Концентрація IL-6
I_{10}	Концентрація IL-10
P	Концентрація PD-1
L	Концентрація PD-L1
Q	Концентрація PD-1-PD-L1

A	Концентрація анти-PD-1
B	Концентрація інгібітора BRAF/MEK

Вважається, що густина уламків мертвих клітин з некрозу або апоптозу є стала. Ми також припускаємо, що густина незрілих дендритних клітин і наївних $CD4^+$ T і $CD8^+$ T-клітин залишаються постійними по всій пухлині тканини. За припущенням (1), проліферація раку клітин і імміграція імунних клітин в пухлину дають підвищення внутрішнього тиску, що призводить до руху клітин. Вважаємо, що всі клітини рухаються з однаковою швидкістю u ; u залежить від простору та часу (в одиницях см/день). Ми також припускаємо, що всі клітини піддаються дисперсії (тобто дифузії), і що всі цитокіни і протипухлинні препарати дифундують всередині пухлини.

Рівняння для DC (D)

Під некротичними раковими клітинами (N_C) ми маємо на увазі ракові клітини, які зазнають процесу некрозу. Вивільнення некротичних ракових клітин HMGB-1 (H) [F13]. Ми моделюємо динаміку N_C і H за такими рівняннями:

$$\begin{aligned}\frac{\partial N_C}{\partial t} + \nabla \cdot (u N_C) &= \delta_{N_C} \Delta N_C + \lambda_{N_C C} C - d_{N_C} N_C, \\ \frac{\partial H}{\partial t} &= \delta_H \Delta H + \lambda_{H N_C} N_C - d_H H,\end{aligned}$$

де $\lambda_{N_C C}$ -- швидкість, з якою ракові клітини стають некротичними; $d_{N_C C}$ -- швидкість, з якою некротичні клітини перетворюються на сміття, а $\lambda_{H N_C}$ -- швидкість, з якою утворюються некротичні клітини HMGB-1.

$\lambda_{N_C C} C$ отримані з живих ракових клітин, $\lambda_{H N_C} N_C$ звільняють від некротичних ракових клітин.

Зауважимо, що хоча на такі молекули, як HMGB-1 або інші білки, може впливати швидкість u , їх коефіцієнти дифузії на кілька порядків більші, ніж коефіцієнти дифузії клітин, тому їхнім швидкісним членом можна знехтувати. Деградація HMGB-1 швидка (0,01/день), і ми припускаємо, що процес некрозу також швидкий. Тоді ми можемо зробити наближення два динамічних рівняння за стаціонарним станом

$$\lambda_{N_C C} C - d_{N_C} N_C = 0, \quad \lambda_{H N_C} N_C - d_H H = 0,$$

так що H є пропорційним C .

Динаміка DC описується рівнянням

$$\frac{\partial D}{\partial t} + \nabla \cdot (uD) = \delta_D \Delta D + \lambda_{DC} D_0 \frac{C}{K_C + C} - d_D D. \quad (2)$$

де δ_D -- коефіцієнт дифузії, а d_D -- швидкість смерті DC.

Рівняння для $CD4^+$ T-клітин (T_1).

Наївні $CD4^+$ T-клітини диференціюються в клітини Th1 (T_1) під IL-12 (I_{12}), тоді як IL-10 і Tregs пригнічують диференціювання наївних $CD4^+$ T-клітин у T_1 -клітини [F11]. Відбувається проліферація активованих T_1 -клітин посиленням IL-2. Вважається, що обидва процеси активації та проліферації T_1 пригнічуються PD-1-PD-L1 (Q) у $1/(1+Q/K_{TQ})$ раз. Отже, T_1 задовольняє таке рівняння:

$$\begin{aligned} \frac{\partial T_1}{\partial t} + \nabla \cdot (uT_1) = & \delta_T \Delta T_1 + (\lambda_{T_1 I_{12}} T_{10} \frac{I_{12}}{K_{I_{12}} + I_{12}} \frac{1}{1 + I_{10}/K_{T_{10}}} \frac{1}{1 + T_r/K_{T_r}} \\ & + \lambda_{T_1 I_2} T_1 \frac{I_2}{K_{I_2} + I_2}) \frac{1}{1 + Q/K_{TQ}} - d_{T_1} T_1. \end{aligned} \quad (3)$$

де T_{10} -- густина наївних $CD4^+$ T -клітин.

Члени у рівнянні (3) мають такий фізичний сенс: $\lambda_{T_1 I_{12}} T_{10} \frac{I_{12}}{K_{I_{12}} + I_{12}}$ -- активація IL-12, $\frac{1}{1 + I_{10}/K_{T_{10}}}$ -- інгібування IL-10, $\frac{1}{1 + T_r/K_{T_r}}$ -- інгібування Tregs, $\lambda_{T_1 I_2} T_1 \frac{I_2}{K_{I_2} + I_2}$ -- IL-2-індукована проліферація $\frac{1}{1 + Q/K_{TQ}}$ -- інгібування PD-1-PD-L1, $d_{T_1} T_1$ -- смерть.

Рівняння для активованих $CD8^+$ T -клітин (T_8).

Неактивні $CD8^+$ T -клітини активуються IL-12, і цьому процесу протистоять IL-10 і Tregs [F11]. IL-2 посилює проліферацію активованих $CD8^+$ T -клітин. Подібно до рівняння для T_1 , T_8 задовольняє рівняння:

$$\begin{aligned} \frac{\partial T_8}{\partial t} + \nabla \cdot (uT_8) = & \delta_T \Delta T_8 + (\lambda_{T_8 I_{12}} T_{80} \frac{I_{12}}{K_{I_{12}} + I_{12}} \frac{1}{1 + I_{10}/K_{T_{10}}} \frac{1}{1 + T_r/K_{T_r}} \\ & + \lambda_{T_8 I_2} T_8 \frac{I_2}{K_{I_2} + I_2}) \frac{1}{1 + Q/K_{TQ}} - d_{T_8} T_8. \end{aligned} \quad (4)$$

де T_{10} -- густина неактивних $CD8^+$ T -клітин.

Рівняння для активованих Tregs (T_r).

Наївні $CD4^+$ T -клітини диференціюються в Tregs (T_r) активацією фактором транскрипції Fox3+. Комплекс PD-1-PD-L1 посилює експресію PTEN, яка призводить до посилення регуляції Fox3+ і, отже, до збільшення виробництва Tregs [F15]. Виробництво T_r також підсилюється TGF- β (T_β). Активовані Tregs є рекрутований в пухлину імуносупресивним цитокином IL-6 (і CCL2). Представляючи хіміоатракціон за допомогою $\nabla \cdot (\chi T_r \nabla I_6)$, ми отримуємо рівняння для T_r :

$$\frac{\partial T_r}{\partial t} + \nabla \cdot (uT_r) = \delta_T \Delta T_r + T_{10} (\lambda_{T_r T_\beta} \frac{T_\beta}{K_{T_\beta} + T_\beta} + \lambda_{T_r Q} \frac{Q}{K_Q + Q}) - d_{T_r} T_r. \quad (5)$$

$\lambda_{T_r T_\beta} \frac{T_\beta}{K_{T_\beta} + T_\beta}$ -- TGF- β -індукована проліферація, $\lambda_{T_r Q} \frac{Q}{K_Q + Q}$ -- просування PD-

1-PD-L1.

Рівняння для активованих MDSC (M).

Пухлина рекрутує макрофаги і «навчає» їх ставати асоційованими з пухлиною макрофагами (TAM), які поведуться як MDSC [F11]. MDSC також хемотаксично притягуються до мікрооточення пухлини IL-6 (і CCL2). Рівняння MDSC має вигляд:

$$\frac{\partial M}{\partial t} + \nabla \cdot (uM) = \delta_T \Delta M + \lambda_M (M_0 - M) \frac{I_6}{K_{I_6} + I_6} - \nabla \cdot (\chi M \nabla I_6) - d_M M, \quad (6)$$

де M_0 -- джерело/приплив макрофагів з крові.

Рівняння для пухлинних клітин (C).

Ракові клітини знищуються клітинами T_1 і T_8 . Інгібітор BRAF/MEK (B), наприклад вемурафеніб/дабрафеніб, використовується для лікування поширеної меланоми. Його механізм дії передбачає вибірковість інгібування мутованої BRAF-кінази, що призводить до зниження передачі сигналів через аномальний шлях RAS-RAF-MEKERK (MAPK). Ми припускаємо, що BRAF/MEK інгібітор пригнічує аномальну проліферацію пухлини клітини в $1/(1 + B/K_{CB})$. Тоді рівняння для C набуває форму:

$$\frac{\partial C}{\partial t} + \nabla \cdot (uC) = \delta_C \Delta C + \lambda_C C \left(1 - \frac{C}{C_M} \right) \frac{1}{1 + B/K_{CB}} - (\eta_1 T_1 C + \eta_8 T_8 C) - d_C C, \quad (7)$$

де η_1, η_8 -- швидкість знищення ракових клітин T_1 і T_8 , а d_C – природна швидкість загибелі ракових клітин.

Рівняння для IL-12 (I_{12}).

$$\frac{\partial I_{12}}{\partial t} = \delta_{I_{12}} \Delta I_{12} + \lambda_{I_{12}} D D \left(1 + \lambda_{I_{12}} B \frac{B}{K_B + B} \right) - d_{I_{12}} I_{12}. \quad (8)$$

Рівняння для IL-2 (I_2).

IL-2 виробляється активованими $CD4^+ T$ -клітинами (T_1). Отже,

$$\frac{\partial I_2}{\partial t} = \delta_{I_2} \Delta I_2 + \lambda_{I_2 T_1} T_1 - d_{I_2} I_2. \quad (9)$$

Рівняння для TGF- β (T_β).

TGF- β продукується пухлинними клітинами:

$$\frac{\partial T_\beta}{\partial t} = \delta_{T_\beta} \Delta T_\beta + \lambda_{T_\beta C} C + \lambda_{T_\beta T_r} T_r + \lambda_{T_\beta M} M - d_{T_\beta} T_\beta. \quad (10)$$

Рівняння для IL-6 (I_6).

IL-6 виробляється раковими клітинами, так що

$$\frac{\partial I_6}{\partial t} = \delta_{I_6} \Delta I_6 + \lambda_{I_6 C} C - d_{I_6} I_6. \quad (11)$$

Рівняння для IL-8. (I_{10})

IL-10 виробляється раковими клітинами та МДСК [F11]. Отже воно задовольняє таке рівняння:

$$\frac{\partial I_{10}}{\partial t} = \delta_{I_{10}} \Delta I_{10} + \lambda_{I_{10} C} C + \lambda_{I_{10} M} M - d_{I_{10}} I_{10}. \quad (12)$$

Рівняння для анти-PD-1 (A).

Ми припускаємо, що анти-PD-1 вводять внутрішньошкірно кожен раз три дні упродовж 30 днів, забезпечує джерело $\hat{A}(t)$ анти-PD-1:

$$\hat{A}(t) = \begin{cases} \gamma_A, & t \leq 30, \\ \gamma_A \frac{33-t}{3}, & 30 \leq t \leq 33, \\ 0, & t > 33. \end{cases}$$

де γ_A – ефективний рівень препарату; хоча для спрощення рівень препарату змінюється між ін'єкціями ми вважаємо його постійним. Препарат A виснажується у процесі блокування ПД-1. Отже,

$$\frac{\partial A}{\partial t} = \delta_A \Delta A + \hat{A}(t) - \mu_{PA} PA - d_A A. \quad (13)$$

Рівняння для інгібітора BRAF/MEK (B).

Ми припускаємо, що вводиться інгібітор BRAF/MEK внутрішньошкірно щодня упродовж 30 днів, забезпечуючи джерело $\hat{B}(t)$ BRAF/MEKi:

$$\hat{B}(t) = \begin{cases} \gamma_B, & t \leq 30, \\ \gamma_B \frac{33-t}{3}, & 30 \leq t \leq 33, \\ 0, & t > 33. \end{cases}$$

Припускаючи, що BRAF/MEKi поглинається C зі швидкістю $\mu_{BC} C \frac{B}{K_B + B}$, ми отримуємо таке рівняння для B :

$$\frac{\partial B}{\partial t} = \delta_B \Delta B + \hat{B}(t) - \mu_{BC} C \frac{B}{K_B + B} - d_B B. \quad (14)$$

Рівняння для швидкості клітин (u).

Ми припускаємо, що частина пухлини складається з позаклітинного матриксу, ЕСМ (приблизно $0,4 \text{ г/см}^3$), ракових клітин (приблизно $C = 0,4 \text{ г/см}^3$) та МДСК (приблизно $M = 0,2 \text{ г/см}^3$). Припускаємо (в розділі оцінка параметрів), що щільність імунних клітин D , T_1 , T_8 і T_r приблизно $4 \cdot 10^{-4}$, $2 \cdot 10^{-3}$, 10^{-3} г/см^3 , відповідно, і, для консистенції, константа у рівнянні (1) має бути 0,6037. Далі вважаємо, що всі клітини мають приблизно однакові об'єми та площі поверхні, щоб коефіцієнти дифузії усіх клітинок були однакові. Додавання правих частин (2)-(7) рівнянь дає

$$0,6039 \cdot \nabla u = \sum_{j=2}^7 [\text{rihgtparts}(j)]. \quad (15)$$

Рівняння вільної межі (R).

Вважаємо, що вільна межа $r = R(t)$ рухається зі швидкістю клітин, так що

$$\frac{dR}{dt} = u[R(t), t]. \quad (16)$$

Межові умови.

Припускається, що наївні $CD4^+$ Т-клітини і неактивні $CD8^+$ Т-клітини, які мігрували з лімфатичних вузлів у мікрооточення пухлини, мають постійні густини $\hat{T}_1$ і $\hat{T}_8$

на межі пухлини, і що T_1 і T_8 активуються ІЛ-12 при проникненні в пухлину. Тоді ми маємо такі умови потоку на межі пухлини:

$$\begin{aligned}\frac{\partial T_1}{\partial r} + \sigma_T(I_{12})(T_1 - \hat{T}_1) &= 0, \\ \frac{\partial T_8}{\partial r} + \sigma_T(I_{12})(T_8 - \hat{T}_8) &= 0 \text{ на } r = R(t),\end{aligned}\quad (17)$$

де $\sigma_T(I_{12}) = \sigma_0 I_{12} / (I_{12} + K_{I_{12}})$.

Межова умова відсутності потоку для всіх інших змінних. Тут негласно передбачається, що рецептори PD-1 і ліганди PD-L1 стають активними тільки після того, як Т-клітини є вже всередині пухлини.

Початкові умови.

Відповідно початкові значення беруться, за якими середня густина ракової клітини ще не досягла стабільного стану. Потім, за рівнянням (1), загальна густина імунних клітин спочатку вище свого стаціонарного стану. Прийmemo (в одиницях г/см^3):

$$\begin{aligned}D &= 2 \cdot 10^{-4}, \quad T_1 = 4 \cdot 10^{-3}, \quad T_8 = 2 \cdot 10^{-3}, \\ T_r &= 3 \cdot 10^{-3}, \quad M = 0,22, \quad C = 0,3774.\end{aligned}\quad (18)$$

Ці початкові умови для клітин задовольняють рівняння (17). Припустимо, що спочатку $B = 0$ і $A = 0$, і візьmemo початкові умови для I_{12} , I_2 , T_β , I_6 , I_{10} і P близькими до їхніх значень усталеного стану, які обчислюються у розділі оцінки параметрів. Один вибір початкової умови наведено таким чином (в одиницях г/см^3):

$$\begin{aligned}I_{12} &= 4 \cdot 10^{-10}, \quad I_2 = 4,74 \cdot 10^{-11}, \quad T_\beta = 2,62 \cdot 10^{-13}, \\ I_6 &= 3,06 \cdot 10^{-11}, \quad I_{10} = 9,66 \cdot 10^{-11}, \quad P = 8,3 \cdot 10^{-10}.\end{aligned}$$

Розглянемо комбіновану терапію для низки значень інгібітора BRAF /MEK та анти-PD-1. Ефективність комбінованої терапії, $z(\gamma_B, \gamma_A)$, визначається за формулою:

$$E(\gamma_A, \gamma_B) = \frac{R_{60}(0,0) - R_{60}(\gamma_A, \gamma_B)}{R_{60}(0,0)},$$

де обчислюється радіус пухлини $R_{60}(\gamma_A, \gamma_B)$ у день 60; $R_{60}(0,0)$ — радіус на 60-й день у контрольному випадку (без наркотиків). Ефективність - це позитивне число та його значення міститься між 0 і 1 (або між 0 і 100%).

Табл. 11.2. Параметри моделі (г/см^3)

Парам.	Опис	Знач.	Розмір.
	Коефіцієнт дифузії		
δ_D	ДК	$8,64 \cdot 10^{-7}$	$\text{см}^2 / \text{доба}$
δ_T	Т-клітин	$8,64 \cdot 10^{-7}$	$\text{см}^2 / \text{доба}$
δ_M	МДСК	$8,64 \cdot 10^{-7}$	$\text{см}^2 / \text{доба}$
δ_C	пухлинних клітин	$8,64 \cdot 10^{-7}$	$\text{см}^2 / \text{доба}$
$\delta_{I_{12}}$	ІЛ-12	$6,05 \cdot 10^{-2}$	$\text{см}^2 / \text{доба}$
δ_{I_2}	ІЛ-2	$9,58 \cdot 10^{-2}$	$\text{см}^2 / \text{доба}$
δ_{T_β}	TGF – β	$8,52 \cdot 10^{-2}$	$\text{см}^2 / \text{доба}$

δ_{I_6}	IL-6	$9,03 \cdot 10^{-2}$	$\text{cm}^2 / \text{доба}$
$\delta_{I_{10}}$	IL-10	$9,11 \cdot 10^{-2}$	$\text{cm}^2 / \text{доба}$
δ_A	анти-PD-1	$7,85 \cdot 10^{-2}$	$\text{cm}^2 / \text{доба}$
δ_B	BRAF/MEKi	$3,16 \cdot 10^{-1}$	$\text{cm}^2 / \text{доба}$
χ	хемоатракції IL-6	10	$\text{cm}^5 / (\text{г д.})$

Табл. 11.3. Параметри моделі (г/см^3)

Парам.	Опис	Знач.	Розмір.
	Швидкість		
σ_0	потоків клітин T_1 і T_8 на межі	1	1/см
λ_{DC}	активації DC пухлинними кл.	4	$\text{g/cm}^3 \text{day}$
$\lambda_{T_1 I_{12}}$	активації T-клітин $CD4^+ T$ IL-12	18,64	1/доба
$\lambda_{T_8 I_{12}}$	активації $CD8^+ T$ -клітин IL-12	16,6	1/доба
$\lambda_{T_1 I_2}$	активації $CD8^+ T$ -клітин IL-12	0,25	1/доба
$\lambda_{T_r T_\beta}$	активації Tregs TGF- β	0,415	1/доба
$\lambda_{T_r Q}$	активації Tregs PD-1-PD-L1	0,083	1/доба
λ_M	активації MDSCs	1,05	1/доба
λ_C	росту ракових клітин	0,616	1/доба
λ_0	росту ракових клітин		
	без гальмування	0,069	1/доба
$\lambda_{I_{12} B}$	сприяння виробництву		
	IL-12 BRAF/MEKi	1	1/доба
$\lambda_{I_{12} D}$	виробництва IL-12 DC	$2,76 \cdot 10^{-6}$	1/доба
$\lambda_{I_2 T_1}$	виробництва IL-2 $CD4^+ T$ -кл.	$2,82 \cdot 10^{-8}$	1/доба
$\lambda_{T_\beta C}$	виробництва TGF- β раковими кл.	$2,18 \cdot 10^{-10}$	1/доба
$\lambda_{T_\beta T_r}$	продукції TGF- β за Tregs	$5,57 \cdot 10^{-9}$	1/доба
$\lambda_{T_\beta M}$	виробництва TGF- β MDSCs	$2,18 \cdot 10^{-10}$	1/доба
$\lambda_{I_6 C}$	виробництва IL-6 раковими кл.	$3,54 \cdot 10^{-10}$	1/доба
$\lambda_{I_{10} C}$	виробництва IL-10 раковими кл.	$9,10 \cdot 10^{-10}$	1/доба
$\lambda_{I_{10} M}$	виробництва IL-10 MDSCs	$1,82 \cdot 10^{-9}$	1/доба
η_1	знищення клітин $CD4^+ T$ -кл.	11,5	$\text{cm}^3 / (\text{г д.})$
η_8	знищення клітин $CD8^+ T$ -кл.	46	$\text{cm}^3 / (\text{г д.})$
μ_{PA}	блокування PD-1 анти-PD-1	$6,04 \cdot 10^6$	$\text{cm}^3 / (\text{г д.})$

Табл. 11.4. Параметри моделі (г/см³)

Параметр	Опис	Знач.	Розмірн.
	Коефіцієнт		
d_D	Смертність ДК	0,1	1/доба
d_{T_1}	Рівень смертності $CD4^+T$ -клітин	0,197	1/доба
d_{T_8}	Рівень смертності $CD8^+T$ -клітин	0,18	1/доба
d_{T_r}	Смертність Tregs	0,2	1/доба
d_M	Рівень смертності MDSCs	0,03	1/доба
d_C	Рівень загибелі пухлинних клітин	0,17	1/доба
	Швидкість		
μ_{BC}	поглинання BRAF/MEKі		
	раковими клітинами	$6,17 \cdot 10^{-10}$	1/доба
$d_{I_{12}}$	деградації IL-12	1,38	1/доба
d_{I_2}	деградації IL-2	2,376	1/доба
d_{T_β}	Швидкість деградації TGF- β	499,066	1/доба
d_{I_6}	деградації IL-6	4,16	1/доба
$d_{I_{10}}$	розпаду IL-10	8,32	1/доба
d_A	деградації анти-PD-1	0,046	1/доба
d_A	деградації BRAF/MEKі	1,66	1/доба
	густина		
D_0	неактивних ДК	$2 \cdot 10^{-5}$	г/см ³
T_{10}	наївних $CD4^+T$ -клітин у пухлині	$4 \cdot 10^{-4}$	г/см ³
T_{80}	наївних $CD8^+T$ -клітин у пухлині	$2 \cdot 10^{-4}$	г/см ³
$\hat{T}_1$	$CD4^+T$ -клітин з лімфатичних вузлів	$4 \cdot 10^{-3}$	г/см ³
$\hat{T}_8$	$CD8^+T$ -клітин з лімфатичних вузлів	$2 \cdot 10^{-3}$	г/см ³
C_M	Несуча здатність ракових клітин	0,8	г/см ³

11.3. Додаток. Оцінка параметрів моделі

Коефіцієнти дифузії.

За [45] коефіцієнт дифузії δ_p та молекулярна маса M_p будь-якої білка p пов'язані формулою $\delta_p = \frac{\text{const}}{\sqrt[3]{M_p}}$.

$$\delta_p = \sqrt[3]{M_G/M_p} \delta_G,$$

де M_G і δ_G – молекулярна маса та коефіцієнт дифузії VEGF відповідно;
 $M_G = 24 \text{ kDa}$, $\delta_G = 8,64 \cdot 10^{-2} \text{ cm}^2/\text{доба}$. $M_{I_{12}} = 70 \text{ kDa}$, $\delta_{I_{12}} = 6,05 \cdot 10^{-2} \text{ cm}^2/\text{доба}$;
 $\delta_{I_2} = 9,58 \cdot 10^{-2} \text{ cm}^2/\text{доба}$; $M_A = 32 \text{ kDa}$, $M_B = 489.93 \text{ Da}$; $\delta_B = 0,316 \text{ cm}^2/\text{доба}$,
 $\delta_{T_\beta} = 8,52 \cdot 10^{-2} \text{ cm}^2/\text{доба}$, $\delta_{I_6} = 9,03 \cdot 10^{-2} \text{ cm}^2/\text{доба}$, $\delta_{I_{10}} = 9,11 \cdot 10^{-2} \text{ cm}^2/\text{доба}$;
 $\delta_A = 0,0785 \text{ cm}^2/\text{д}$.

Коефіцієнт дифузії кисню у позаклітинному матриксі (ЕМК) знаходиться в межах $(0,7 \div 2) \cdot 10^{-5} \text{ cm}^2/\text{с}$, отримуємо для $\delta_w = 0,8 \text{ cm}^2/\text{доба}$.

Рівняння (2).

Зі стаціонарного стану рівняння (2) отримуємо

$$\lambda_{DC} D_0 \frac{C}{K_C + C} \frac{1}{1 + G/K_{DG}} = d_D D,$$

де $d_D = 0,1/\text{доба}$, $C = K_C = 0,4 \text{ g/cm}^3$, $D = K_D = 4 \cdot 10^{-4} \text{ g/cm}^3$, $D_0 = 2 \cdot 10^{-5} \text{ g/cm}^3$. Ми припускаємо, що $K_{DG} = 4K_G$, $K_G = 7 \cdot 10^{-8} \text{ g/cm}^3$.

Використовуючи рівняння стаціонарного стану, ми ігнорували той факт, що об'єм пухлини збільшується, що спричиняє розведення середньої густини. Нам потрібно це розведення компенсувати збільшенням терміну виробництва. Прийmemo $\lambda_{DC} = 6/\text{день}$. При призначенні хіміотерапевтичного препарату, приймаємо $\varepsilon_{DA} = 10^8 \text{ cm}^3/\text{г}$.

Рівняння (3).

Вважаємо, що $K_{TG} = 4K_G$, $K_G = 7 \cdot 10^{-8} \text{ g/cm}^3$. Покладемо праву частину рівняння рівною нулю (стаціонарний стан). Матимемо

$$\lambda_{T_8 I_{12}} T_{80} \cdot \frac{1}{2} \cdot \frac{4}{5} - d_{T_8} T_8 = 0,$$

де $d_{T_8} = 0,18/\text{доба}$, $T_{80} = 2 \cdot 10^{-4} \text{ g/cm}^3$, $T_8 = K_{T_8} = 10^{-3} \text{ g/cm}^3$.

Отже, $\lambda_{T_8 I_{12}} T_{80} = 2,25/\text{доба}$.

Через ефект розведення (як при оцінці $\lambda_D C$), нам потрібно збільшити $\lambda_{T_8 I_{12}}$. Але, оскільки межова умова для T_8 приносить нові клітини в пухлинну тканину, яка була проігнорована в рівнянні стаціонарного стану, нам потрібно, в той же час також зменшити $\lambda_{T_8 I_{12}}$. Ми припускаємо, що обидва ефекти компенсуються.

Рівняння (4).

$d_E = 0,69/\text{доба}$, $E_M = 5 \cdot 10^{-3} \text{ g/cm}^3$, $K_E = 2,5 \cdot 10^{-3} \text{ g/cm}^3$, $\hat{E} = 5 \cdot 10^{-3} \text{ g/cm}^3$,
 $G_0 = 3,65 \cdot 10^{-10}$. Беремо $E_M = 2K_E = 5 \cdot 10^{-3} \text{ g/cm}^3$. Зі стаціонарного стану рівняння (17) (ми

можемо ігнорувати хемотаксичний член, оскільки немає потоку G від межі), ми отримуємо $\lambda_{EG} = 2d_E/(K_G - G_0) = 1,98 \cdot 10^7 \text{ cm}^3/(\text{g доба})$.

Тут, як і у випадку $\lambda_{T_8 I_{12}}$, потрібно розглянути вплив на λ_{EG} через розведення, а також ефект припливу ендотеліальних клітин, стимульованих VEGF. Ми припускаємо, що приплив ендотеліальних клітин більш домінуючий, і збільшити в 1,28 рази, так що $\lambda_{EG} = 2,28 \cdot 10^7 \text{ cm}^3/(\text{g доба})$.

Рівняння (5).

Беремо $d_C = 0,17 \text{ 1/доба}$, $C_M = 0,8 \text{ g/cm}^3$ і $\lambda_{CW} = 1,6 \text{ 1/доба}$. У стаціонарному стані контрольного випадку (без протипухлинних препаратів) ми припускаємо, що $c \approx 0,4 \text{ g/cm}^3$, $W = W_0 - K_W = 1,69 \cdot 10^{-4} \text{ g/cm}^3$. У контрольному випадку зі стаціонарного стану рівняння ((182)) маємо

$$0,5\lambda_{CW}K_W/W_0 - \eta_8 K_{T_8} - d_C = 0,$$

де $K_{T_8} = 10^{-3} \text{ g/cm}^3$, $\eta_8 = (\lambda_C K_W / (2W_0) - d_C) / K_{T_8} = 120,75 \text{ cm}^3/(\text{г. доба})$. У контрольному випадку, включаючи ефект адвекції і те, що пухлина росте, нам потрібно збільшити швидкість росту ракові клітини; ми приймаємо $\lambda_{CW} = 1,9 \text{ /доба}$. Коли вводять хіміотерапевтичний препарат і анти-VEGF, ми приймаємо $\mu_{CA} = 10^8 \text{ sm}^3/\text{g}$, $K_{AB} = 10^{-9} \text{ g/cm}^3$.

Рівняння (7).

Зі стаціонарного стану рівняння (7) отримуємо $\lambda_{WE}E - d_W W = 0$, де $\lambda_{WE} = 7 \cdot 10^{-2} \text{ /доба}$, $W = k_W = 1,69 \cdot 10^{-4} \text{ g/cm}^3$, $E = K_E = 2,5 \cdot 10^{-3} \text{ g/cm}^3$, $G = K_G = 7 \cdot 10^{-7} \text{ g/cm}^3$, $C = K_C = 0,4 \text{ g/cm}^3$. Отже, $d_W = \lambda_{WE}E/W = 1,04 \text{ /доба}$. Зверніть увагу, оскільки ми вже враховуючи ефект розведення для E , ми не включаємо його в рівняння для W .

Рівняння (8).

Зі стаціонарного стану рівняння отримуємо $\lambda_{GW}C - d_G G = 0$, $d_G = 12,6 \text{ 1/доба}$, $G + K_G = 7 \cdot 10^{-8} \text{ g/cm}^3$, $C = K_C = 0,4 \text{ g/cm}^3$. Отже, $\lambda_{GW} = 2,21 \cdot 10^{-6} \text{ /доба}$.

Бібліографія

До розділу 7.

7.1. Alzahrani T., Eftimie R., Trucu D. Multiscale Modelling of Cancer Response to Oncolytic Viral Therapy, *Mathematical Biosciences*, 2019, 22 p.

7.2.. Andasari V., Gerisch A., Chaplain M. Mathematical modeling of cancer cell invasion of tissue: biological insight from mathematical analysis and computational simulation, *J. Math. Biol.*, 63 (1), 2011, pp. 141--171.

7.3. Burden T., Ernstberger J., Fister K. Optimal control applied to immunotherapy, *Disc Cont Dyn Syst Ser B*, 4(1), 2004, pp. 135--146.

7.4. Britton N., Essential Mathematical Biology, Springer Verlag, 2003.

7.5. Castiglione F., Piccoli B. Optimal control in a model of dendritic cell transfection cancer immunotherapy, *Bull Math Biol*, 68(2), 2006, pp. 255--274.

7.6. Cesca M., Morosi L., Berndt A. Bevacizumab-induced inhibition of angiogenesis promotes a more homogeneous intratumoral distribution of paclitaxel, improving the antitumor response, *Molecular Cancer Therapeutics*, 15 (1), 2016, pp.125--135.

7.7. Chaplain M., Lolas G. Mathematical modelling of cancer cell invasion of tissue: the role of the urokinase plasminogen activation system, *Math. Model. Meth. Appl. Sci.*, 15 (11), 2005, pp. 1685--1734.

7.7. Chanmee T., Ontong P., Konno K. Tumor-associated macrophages as major players in the tumor microenvironment, *Cancers*, 6 (3), 2014, pp. 1670--1690.

7.7. Kirschner D., Panetta J. Modeling immunotherapy of the tumor-immune interaction, *J Math Biol.*, 37(3), 1998, pp. 235--252.

7.8. Kuznetsov V., Makalkin I., Taylor M., Perelson A. Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis, *Bull Math Biol.*, 56(2), 1994, pp. 295--321.

7.9. Lin A. A model of tumor and lymphocyte interactions, *Disc Cont Dyn Syst Ser B*, 4(1), 2004, pp. 241--266.

7.10. Mollard S., Ciccolini J., Barbolosi D. Model driven optimization of antiangiogenics+ cytotoxics combination, *Oncotarget*, 8 (14), 2017, pp. 23087--97.

7.13. Muppidi M., George S. Immune checkpoint inhibitors in renal cell carcinoma. *J Targeted Ther Cancer*, 4, 2015, pp. 47--52.

7.14. Palucka J., Banchereau J. Cancer immunotherapy via dendritic cells, *Nat Rev Cancer*, 12 (4), 2012, pp. 265--277.

7.15. de Pillis L., Radunskaya A. The dynamics of an optimally controlled tumor model: a case study, *Math Comput Model*, 37, 2003, pp. 1221--1244.

7.16. de Pillis L., Gu W., Radunskaya A. Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations, *J Theor Biol.*, 238(4), 2006, pp.841--862.

7.17. de Pillis L., Fister K., Gu W. Chemotherapy for tumors: an analysis of the dynamics and a study of quadratic and linear optimal controls, *Math Biosci.*, 209, 2007, pp. 292--315.

7.17. Trucu D., Lin P., Chaplain M. A multiscale moving boundary model arising in cancer invasion, *Multiscale Model. Simul.*, 11 (1), 2013, pp. 309--335.

7.17. Wodarz D. Computational modelling approaches to the dynamics of oncolytic viruses, *Wiley Interdiscip. Rev. Syst. Biol.*, 8, 2016, pp. 242--252.

До розділу 8.

8.1. Auffinger D., Ahmed A., Lesniak M. Oncolytic virotherapy for malignant glioma: Translating laboratory insights into clinical practice, *Front. Oncol.*, 3, 2013, pp. 1--32.

8.2. Friedman A., Tian J., Chioca E. Glioma virotherapy: Effects of innate immune suppression and increased viral replication capacity, *Cancer Res.*, 66, 2006, pp. 2314–2317.

8.3. Ratajczyk E., Ledzewicz U., Friedman A. The role of $\text{tnf-}\alpha$ inhibitor in glioma virotherapy: a mathematical model, *Mathematical biosciences and engineering*, V. 14, No 1, 2017, pp. 305–317.

8.4. Schattler H., Ledzewicz U. Optimal Control for Mathematical Models of Cancer Therapies, Springer Publishing Co., New York, USA, 2015.

8.5. Wollmann G., Ozduman K., van den Pol A. Oncolytic virus therapy for glioblastoma multiforme: Concepts and candidates, *Cancer J.*, 18, 2012, pp. 69–81.

8.6. Wu J., Byrne H., Kirn D. Modeling and analysis of a virus that replicates selectively in tumor cells, *Bull. Math. Biol.*, 63, 2001, pp. 731–767.

8.7. Dmitrieva N., Yu L., Viapiano M. Chondroitinase ABC I-mediated enhancement of oncolytic virus spread and antitumor efficacy. *Clin Cancer Res.*, 17(6), 2011, pp. 1362–1377.

8.7. Emre Y., Imhof B. Matricellular protein CCN1/CYR61: a new player in inflammation and leukocyte trafficking. *Sem Immunopathol*, 2014, Springer 36, pp. 253–257.

8.7. Hao W., Friedman A. The LDL–HDL profile determines the risk of atherosclerosis: a mathematical model. *PloS One*, 9(3), 2014, e90497.

8.8. Haseley A., Boone S., Wojton J. Extracellular matrix protein CCN1 limits oncolytic efficacy in glioma. *Cancer Res.*, 72(6), 2012, pp. 1353–136R

8.9. Jacobsen K., Russell L., Friedman A. Effects of CCN1 and Macrophage Content on Glioma Virotherapy: A Mathematical Model, *Bull Math Biol.*, 2015, 29 P.

8.10. Kim Y., Lee H., Dmitrieva N. Friedman A. Choindroitinase ABC I-mediated enhancement of oncolytic virus spread and anti tumor efficacy: a mathematical model. *PloS One*, 9(7), 2014, pp. e102497.

8.13. Msaouel P., Dispenzieri A., Galanis E. Clinical testing of engineered oncolytic measles virus strains in the treatment of cancer: an overview. *Curr Opin Mol Ther.*, 11(1), 2009, 43.

8.14. Parney I., Waldron J., Parsa A. Flow cytometry and in vitro analysis of human glioma-associated macrophages. *J Neurosurg*, 110(3), 2009, p. 57R

8.15. Thorne A., Meisen W., Russell L. Role of cysteine-rich 61 protein (CCN1) in macrophage-mediated oncolytic herpes simplex virus clearance. *Mol Ther.*, 2(9), 2014, pp. 1678–1687.

8.16. Yoo J., Haseley A., Bratasz A. Antitumor efficacy of 34.5 ENVE: a transcriptionally retargeted and “Vstat120”-expressing oncolytic virus. *Mol Ther.*, 20(2), 2012, pp. 287–297.

До розділу 9.

9.1. Andasari V., Gerisch A., Lolas G. Mathematical modeling of cancer cell invasion of tissue: biological insight from mathematical analysis and computational simulation. *J Math Biol.*, 63(1), 2011, pp. 141–171.

9.2. Anderson A., Chaplain M., Newman E. Mathematical modelling of tumour invasion and metastasis. *J Theor Med.*, 2, 2000, pp. 129–154.

9.3. Chaplain M., Lachowicz M., Szymanska Z. Mathematical modelling of cancer invasion: the importance of cell–cell adhesion and cell-matrix adhesion. *Math Model Methods Appl Sci.*, 21(4), 2011, pp. 719–743.

9.4. Cristini V., Lowengrub J., Nie Q. Nonlinear simulation of tumour growth. *J Math Biol.*, 46, 2003, pp. 191–224.

9.5. Domschke P., Trucu D., Gerisch A., Chaplain M. Mathematical modelling of cancer invasion: implications of cell adhesion variability for tumour infiltrative growth patterns. *J Theor Biol.*, 361, 2014, pp. 41–60.

9.6. Peng L., Trucu D., Lin P. A Multiscale Mathematical Model of Tumour Invasive, Growth, *Bull Math Biol.*, 79 (3), 2017, pp. 309–335.

9.7. Preziosi L., Tosin A. Multiphase modelling of tumour growth and extracellular matrix interaction: mathematical tools and applications. *J Math Biol.*, 58, 2009, pp. 625–656.

9.7. Trucu D., Lin P., Chaplain M., Wang Y. A multiscale moving boundary model arising in cancer invasion. *Multiscale Model Simul.*, 11(1), 2013, pp. 309–335.

До розділу 10.

10.1. Ascierto P., Kirkwood J., Grob J. The role of braf v600 mutation in melanoma. *J Transl Med.*, 10(85), 2012, pp. 1–7.

10.2 Eroglu Z., Ribas A. Combination therapy with braf and mek inhibitors for melanoma: latest evidence and place in therapy. *Ther Adv Med Oncol.*, 8(1), 2016, pp. 48–56.

10.3 Friedman A., Hao W., The role of exosomes in pancreatic cancer microenvironment, *Bulletin of Mathematical Biology*, 2017, pp. 1–23.

10.4. Haabeth O., Tveita A., Lørvik K. How do cd4(+) t cells detect and eliminate tumor cells that either lack or express mhc class ii molecules? *Front Immunol.*, 5(174), 2014, pp. 1–13.

10.5. Hargadon K., Bishop J., Brandt J. Melanoma-derived factors alter the maturation and activation of differentiated tissue-resident dendritic cells. *Immunol Cell Biol.*, 2016, 94(1), pp. 24–37.

10.6. Lai X., Friedman A. Mathematical modeling in scheduling cancer treatment with combination of VEGF inhibitor and chemotherapy drugs. *Journal of Theoretical Biology*, 15, 2018, 22 p.

10.7. Lai X., Friedman A. Combination therapy of cancer with braf inhibitor and immune checkpoint inhibitor: A mathematical model, *BMC System Biology*, 11 (70), 2017, pp. 1–17.

10.7. Menzies A., Long G. Recent advances in melanoma systemic therapy. braf inhibitors, ctla4 antibodies and beyond. *Eur J Cancer*, 49(15), 2013, pp. 3229–241.

10.7. Munn D., Mellor A. Ido in the tumor microenvironment: Inflammation, counter-regulation, and tolerance. *Trends in Immunology*, 37(3), 2016, pp. 193–207.

10.8. Perrot C., Javelaud D., Mauviel A. Insights into the transforming growth factor-beta signaling pathway in cutaneous melanoma. *Ann Dermatol.*, 25(2), 2013, pp. 135–44.

10.9. Sims G., Rowe D., Herbst R. Hmgb1 and rage in inflammation and cancer. *Annu Rev Immunol.*, 28, 2010, pp. 367–87.

10.12; Sugiura K., Sugiura M. Immune checkpoint blockade therapy: Merits and demerits. *J Clin Exp Dermatol Res.*, 7(372), 2016.

10.13. Umansky V., Blattner C., Gebhardt C., Utikal J. The role of myeloid-derived suppressor cells (mdsc) in cancer progression. *Vaccines*, 4(36), 2016, pp. 1–16.