Міністерство освіти і науки України
Національний університет водного господарства
та природокористування

П. М. Грицюк, О. І. Джоші, О. М. Гладка

ОСНОВИ ТЕОРІЇ СИСТЕМ І УПРАВЛІННЯ

Навчальний посібник

Рівне – 2021

УДК 330.46:005.1(075.8)
Г85

Рецензенти:

Вовк В. М., д-р економ. наук, проф., завідувач кафедри економічної кібернетики Львівського національного університету імені Івана Франка;

Мальчик М. В., д-р економ. наук, проф., завідувач кафедри маркетингу Національного університету водного господарства та природокористування, м. Рівне;

Кристопчук М. Є., канд. техн. наук, завідувач кафедри транспортних технологій і технічного сервісу Національного університету водного господарства та природокористування, м. Рівне.

Рекомендовано вченою радою Національного університету водного господарства та природокористування
Протокол № 3 від 26 березня 2021 р.

Грицюк П. М., Джоші О. І., Гладка О. М.
Г85 Основи теорії систем і управління : навч. посіб. Рівне : НУВГП, 2021. – 272 с.
ISBN 978-966-327-495-9

В посібнику розглядаються засади теорії систем, методологія системних досліджень, методи представлення даних, інформаційні аспекти вивчення систем, методи аналізу і прогнозування часових рядів, статистичні методи обробки даних, системний аналіз транспортних мереж, умови стійкості систем, моделювання та прогнозування систем методами нелінійної динаміки, методи прийняття рішень за умов невизначеності та підходи до управління системами. Мета посібника – формування компетентностей із застосування системної методології для аналізу, моделювання та прогнозування складних об'єктів, які функціонують у відповідності до множини суперечливих критеріїв і цілей, за наявності суттєвих ризиків, невизначеностей та обмеженого обсягу інформації.

Посібник призначено для здобувачів вищої освіти економічних та технічних спеціальностей, для фахівців відповідних напрямків, управлінців та системних аналітиків.

УДК 330.46:005.1(075.8)

ISBN 978-966-327-495-9

ЗМІСТ

ВСТУП

Сьогодні ми є свідками тісної інтеграції в усіх сферах людської діяльності. Сучасні економічні, політичні, соціальні та інформаційні процеси активно взаємодіють між собою; більш взаємозалежними стають держава і суспільство, виробництво і наука, культура та побутова сфера. Більшість сучасних фірм, організацій, підприємств, корпорацій інтегровані в системи міжнаціональних економічних зв'язків, у транснаціональні компанії, в інформаційні системи, що обслуговують світовий ринок.

Можна говорити про настання епохи наукового, системно-міждисциплінарного підходу до проблем науки, освіти, техніки і технологій, епохи, що концентрує увагу не тільки на матеріально-енергетичних, але і на системно-міждисциплінарних аспектах, побудові та дослідженні системно-інформаційної картини світу, про настання епохи системних парадигм.

За таких умов при вивченні економічних процесів недостатнім є застосування лише традиційних аналітичних методів дослідження, необхідні цілісні, комплексні та всебічні підходи, що акцентують увагу не тільки на певному економічному об'єкті, а й на дослідженні навколишнього середовища, в якому він функціонує. Одним із таких методів є системний підхід, що розглядає економіку як складну цілісну систему в різних аспектах: як сукупність елементів різних рівнів агрегування (макрорівень, галузі та сектори економіки, мікрорівень), у розрізі сфер діяльності (виробнича і невиробнича) та функцій (маркетинг, фінанси, аудит тощо) [30, С. 3].

Для фахівців з інформаційних систем та технологій важливим є вирішення проблеми ефективного управління великими інформаційними системами, до складу яких входять сотні комп'ютерів, йотабайти інформації і різноманітного системного та прикладного програмного забезпечення. Для таких складних систем використання класичних аналітичних методів є неможливим, а натурні експерименти дуже обмежені. Тому як основні для дослідження і проектування таких систем використовуються методи системного аналізу, а експерименти реалізуються в комп'ютерному варіанті шляхом побудови та використання системних імітаційних моделей [15, С. 9–10].

Головною метою вивчення дисципліни «Основи теорії систем і управління» є формування системного мислення, усвідомлення необхідності застосування системного підходу до завдань управління та прийняття рішень, до дослідження складних явищ і процесів у соціально-економічних та інформаційних системах.

В рамках дисципліни «Основи теорії систем і управління» вивчаються основні поняття та методології системного аналізу складних взаємопов'язаних об'єктів різної природи, які функціонують у відповідності до множини суперечливих критеріїв і цілей, за наявності суттєвих ризиків та невизначеностей. В даному посібнику розглядаються засади теорії систем, методологія системних досліджень, методи представлення даних, інформаційні аспекти вивчення систем, методи аналізу і прогнозування часових рядів, статистичні методи обробки даних, системний аналіз транспортних мереж, умови стійкості систем, моделювання та прогнозування систем методами нелінійної динаміки, методи прийняття рішень за умов невизначеності та підходи до управління системами.

Мета пропонованого посібника – сформувати фахові компетентності із застосування системної методології для аналізу, моделювання та прогнозування складних об'єктів, побудови комп'ютерних інформаційних систем; розвинути практичні навички логіко-фізичного моделювання та проектування інформаційних систем; ознайомити здобувача з методологією дослідження систем за умов обмеженого обсягу інформації.

Теоретичним фундаментом для вивчення дисципліни «Основи теорії систем і управління» є вища математика, дискретний аналіз, теорія ймовірностей і математична статистика, економічна кібернетика, дослідження операцій і математичне програмування, теорія графів тощо. Технічні засоби системного аналізу і управління – сучасна комп'ютерна техніка та інформаційні системи.

Посібник призначений для здобувачів вищої освіти економічних та технічних спеціальностей, для фахівців відповідних напрямків, управлінців та системних аналітиків.

1. ЗАСАДИ ТЕОРІЇ СИСТЕМ

1.1. Історія розвитку теорії систем і системного аналізу, предмет та мета системного аналізу, системне мислення

Історія розвитку загальної теорії систем пов'язана із розвитком двох понять: системності та керування (зокрема, прийняття рішень), які усвідомлено чи неусвідомлено застосовували ще з давніх-давен.

Слово «система» використовувалося ще в Стародавній Греції понад 2000 років тому. Вчені (Арістотель, Демокріт, Платон та ін.) розглядали складні тіла, процеси чи міфи світобудови як системи, складені з різних елементів, наприклад, атомів, метафор. Розвиток астрономії (Коперник, Галілей, Ньютон та ін.) і формування геліоцентричної системи світу дозволило перейти до категорій типу «річ і властивості», «ціле і частина», «субстанція і атрибути», «подібність і відмінність» тощо. При цьому постулат (Аристотеля): «Важливість цілого вище важливості його складових» змінився на постулат (Галілея): «Ціле пояснюється властивостями його складових».

В епоху зародження основ системності розглядалися, в основному, системи фізичного чи філософського походження, а її розвиток відбувався під впливом різних філософських поглядів, теорій про структуру пізнання і можливості передбачення (Бекон, Гегель, Ламберт, Кант, Фіхте та ін.).

Проте, *системний аналіз*, чиї основи є досить давніми, – порівняно молода наука (порівняна за віком, наприклад, з кібернетикою), що виникла як окрема наукова дисципліна внаслідок *спроб практичного застосування результатів теорії систем до задач керування складними системами*.

Натуралісти XIX–XX ст. (Богданов, Берталанфі, Вінер, Ешбі, Цвіккі та ін.) не тільки актуалізували роль модельного мислення і моделей в природознавстві, а й сформували основні системоутворюючі принципи, принципи системності наукового пізнання, «поєднали» теорію відкритих систем, філософські принципи і досягнення природознавства. Свій сучасний розвиток теорія систем і системний аналіз отримали під впливом досягнень як класичних областей науки (математика, фізика, хімія, біологія, історія та ін.), так і нових некласичних (синергетика, інформатика, когнітологія, теорії нелінійної динаміки і динамічного хаосу, теорія катастроф, нейроматематика, нейроінформатика тощо).

У [9, С. 7–8; 15, С. 13–23] визначено такі основні історичні віхи, які супроводжували наукове становлення теорії систем.

I. Вперше питання про науковий підхід до керування складними системами було поставлене М. А. Ампером у роботі «Дослідження філософії наук, або аналітичний виклад класифікації всіх людських знань» (част. I – 1834 р., II – 1843 р.), де виділено *науку про керування державою, що названа кібернетикою*.

II. Польський вчений Броніслав Трентовський, професор Фрайбургського університету, у 1843 р. видав у Познані книгу «Ставлення філософії до кібернетики як до мистецтва керування народом». Особливістю праці Трентовського було відображення наукових основ практичної діяльності керівника – «гібернета». Грецьке слова χοβερνω (гіберно) було добре відомим ще в XIX-му сторіччі – адміністративна одиниця, населена людьми. В ширшому сенсі – об'єкт керування, до складу якого входять люди, а χοβερνετ (гібернет) – особа, яка керує ресурсами та людьми на певній території, яка повиннна вміти, виходячи із загального блага, примиряти деякі суперечності, а інші – загострювати, скеровуючи розвиток до потрібної мети. За Трентовським *дійсно ефективне керування повинно враховувати всі внутрішні та зовнішні чинники, що впливають на об'єкт керування, а головна складність його реалізації пов'язана зі складністю поведінки людей.* Б. Трентовський далеко просунувся в розумінні та усвідомленні системності людських колективів, складності керування людьми.

III. Наступний етап у вивченні системності пов'язаний з прізвищем О. О. Богданова (справжнє прізвище – Малиновський), який протягом 1911–1925 рр. видав 3 томи книги «Всеобщая организационная наука (тектология)». За Богдановим у суспільстві (і в біологічних системах) існує функціональна сторона (прагнення швидко адаптуватися) і консервативна (архітектурна схема організації). Лише активне використання зовнішнього середовища забезпечує збереженість системи. Здійснюючи *позитивну селекцію*, система за рахунок зовнішнього середовища збільшує кількість внутрішніх зв’язків, підвищує свою складність, покращуючи тим самим ефективність свого функціонування. *Негативна селекція* видаляє всі вибухонебезпечні джерела, долаючи внутрішній антагонізм організації, підвищує її однорідність, порядок в ній, систематизацію, структурну стійкість. Але разом з тим негативна селекція зменшує функціональну ефективність організації.

IV. Відчутний вплив на усвідомлення деяких аспектів поняття системності та особливо керування мають роботи Н. Вінера. Його праця «Кібернетика», що вийшла з друку у 1948 р. визначає ***кібернетику як «науку про управління та зв’язок в тваринах та машинах»***. Пізніше Вінер почав аналізувати процеси в людському суспільстві з точки зору кібернетики.

V. Виникнення ***загальної теорії систем*** (ЗТС) пов’язують з іменем австрійського фізіолога Людвіга фон Берталанфі, який у 20–30-і роки XX століття займався питаннями системного підходу при вивченні живих організмів, розвиваючи ідею про необхідність цілісного підходу в біології та фізіології. У 1956 р. він організував наукове товариство з досліджень у області ЗТС, що видавало щорічні збірники наукових праць, в яких системний підхід розглядався як універсальна концепція, що об’єднує інтереси різноманітних наук. У 1962–1968 рр. Л. фон Берталанфі включав в ЗТС багато наук: кібернетику, теорію інформації, теорію рішень, топологію, факторний аналіз, теорію множин, теорію мереж, теорію автоматів, теорію масового обслуговування, теорію графів.

Системний підхід виник як реакція на розвиток аналітичних підходів в науці, серед яких слід виділити: логічний позитивізм, аналітичну дедукцію, редукціонізм, казуальну (причинну) логіку, індуктивний підхід [15, С. 19–23].

Логічний позитивізм стверджує, що існує «об'єктивна» реальність, яка є незалежною та неспотвореною нашими суб'єктивними інтерпретаціями світу. Однак факти багатовимірні і можуть інтерпретуватися по-різному: кожна група вчених надаватиме особливе значення такому підходу до розв'язання складних проблем, який є найсумісніщим з її філософією та методологією.

Аналітична дедукція та редукціоністська логіка стверджують, що ціле найкраще можна пояснити шляхом пояснення його частин, тобто редукціоніст розв'язує складну проблему шляхом розбиття її на складові та окремого дослідження кожної з них, що приводить до розвитку спеціалізованих дисциплін з певними сферами дослідження та впливу. Проте, виникає множинність в підходах, вчені спілкуються в межах своїх дисциплін, не розуміють наукову мову (тезаурус предметної області) один одного і не в стані оперувати з системними проблемами.

Редукціонізм є позитивним явищем у тому сенсі, що він забезпечує концептуальну основу, засоби і процедури для ідентифікації та вивчення важливих чинників, що впливають на проблему. Однак дедуктивні методи погано працюють, якщо є багато пов'язаних між собою чинників або вони неусвідомлені як чинники.

У більшості випадків наше мислення ґрунтується на концепції причинності, монолітної *казуальної (причинної) логіки*. Згідно з детерміністською концепцією, спостереження (колишні стани системи) разом із законами природи визначають її майбутній стан.

Індуктивний погляд ґрунтується на узагальненні окремих спостережень, формуванні теорії, базуючись на досвіді та знаннях.

Дедуктивний метод – це метод наукового дослідження, який полягає у висуненні гіпотез про причини досліджуваних явищ і виведенні висновків з цих гіпотез шляхом дедукції. Якщо одержані наслідки відповідають усім фактам, що дані у гіпотезі, то ця гіпотеза вважається достовірним знанням. Дедуктивний метод – це своєрідний різновид аксіоматичного методу, коли в природознавстві у ролі аксіом виступають фундаментальні закони природи чи природничо-наукові гіпотези.

Системний підхід синтезує індуктивний та дедуктивний спосіб мислення з залученням інтуїтивних підходів (натхнення, образне мислення тощо).

Предметом системного аналізу, загалом, є розв’язання складних проблем. Поняття «проблема» має людські риси сприйняття, що породжує наступні труднощі:

- неясність розуміння проблеми;
- складнощі постановки проблеми на віддалену перспективу;
- складність класифікації проблеми і, як наслідок, вибір неадекватних засобів її розв’язання;
- спотворена оцінка проблеми (близькі, але дрібні проблеми затуляють великі, але віддалені);
- неправильна оцінка значимості проблеми внаслідок вузькопрофесійної чи некомпетентної точки зору;
- змішування цілей, які необхідно досягнути, з засобами їх досягнення.

Окрім того, проблеми розрізняються за ступенем їх структурованості:

- добре структуровані та сформульовані кількісно;
- слабко-структуровані, в яких зустрічаються як кількісні, так і якісні оцінки;
- неструктуровані, якісні проблеми.

Перший тип проблем не потребує системного аналізу, оскільки існує потужний апарат математичного моделювання та строгі числові методи розв’язання. Основною областю застосування методів системного аналізу є слабко-структуровані проблеми, а для розв’язання неструктурованих проблем здебільшого застосовуються евристичні методи та методи штучного інтелекту.

Метою застосування системного аналізу до конкретної проблеми є підвищення ступеня обґрунтованості рішення, що приймається, та управління системою.

Ця наука, як і будь-яка інша, має на меті дослідження нових зв’язків і відносин об’єктів та явищ. Але, при цьому, основною проблемою є дослідження зв'язків і відносин таким чином, щоб об'єкти, які вивчаються, стали більш керованими, а «відкритий» в результаті дослідження механізм взаємодії цих об'єктів – якнайбільш застосовним до інших аналогічних об'єктів чи явищ. Завдання і принципи системного підходу не залежать від природи досліджуваних процесів.

Системний аналіз відрізняється від інших методів дослідження тим, що:

- враховує принципову складність об'єкта, що досліджується; бере до уваги розгалужені та стійкі взаємозв'язки його з оточенням; враховує неможливість спостереження низки властивостей об'єкта та оточуючого середовища;
- реальні явища чи процеси, їх властивості та зв'язки переводяться в абстрактні категорії теорії систем;
- ґрунтуючись на відомих властивостях складних систем дозволяє виявити нові конкретні властивості та взаємозв'язки конкретного об'єкта дослідження;
- на відміну від інших методів, в яких об'єкти точно визначені, включає як один з важливих етапів визначення об'єкта, його ідентифікацію чи конструювання;
- орієнтується не на розв'язання «правильно сформульованих» задач, а на створення правильної постановки задачі та вибір відповідних методів її розв'язання;
- пропонує шлях, яким можна перетворити складну проблему на простішу; шукає відповідь на запитання: яким чином складну не лише для розв'язання, але й для розуміння, проблему перетворити на послідовність задач, для яких існують методи розв'язання.

Системний аналіз «надає в користування» різноманітним наукам такі системні методи і процедури: абстрагування і конкретизація; аналіз і синтез; індукція і дедукція; формалізація і конкретизація; композиція і декомпозиція; лінеаризація і виокремлення нелінійних складових; структурування і реструктурування; макетування; реінжиніринг; алгоритмізація; моделювання і експеримент; програмне керування і регулювання; розпізнавання і ідентифікація; кластеризація і класифікація; експертне оцінювання і тестування; верифікація та ін.

Потреба в системному аналізі виникає в наступних ситуаціях:

- розв'язується нова проблема, що формулюється, визначається за допомогою системного аналізу;
- розв'язання проблеми передбачає координацію цілей з множиною засобів їх досягнення;
- проблема має розгалужені зв'язки, що викликають віддалені наслідки, і прийняття рішення потребує врахування сукупної ефективності та повних затрат;
- існують варіанти розв'язання проблеми або досягнення взаємопов'язаного комплексу цілей, які важко порівняти;
- створюється нова складна система;
- здійснюється вдосконалення, реконструювання виробництва, реінженерія бізнес-процесів;
- при створенні інформаційних систем та комп'ютеризованих систем керування;
- коли важливі рішення мають ухвалюватися за наявності невизначеності і ризику та (або) на достатньо віддалену перспективу [9, С. 9–13; 15, С. 23–28].

В природі і в суспільстві існують такі основні типи **ресурсів**: ***речовина*** – найкраще вивчений ресурс, який, в основному, поданий таблицею Д. І. Менделєєва, і є відображенням сталості матерії в природі, мірою однорідності матерії; ***енергія*** – не повністю вивчений ресурс (наприклад, ми не володіємо керованою термоядерною реакцією чи електричною енергією блискавки), який є відображенням мінливості матерії, переходів з одного виду в інший, мірою необоротності матерії; ***інформація*** – мало вивчений ресурс, який є відображенням порядку, структурованості матерії (і соціуму); ***людина*** – носій інтелекту вищого рівня, який є в економічному, соціальному та гуманітарному сенсі найважливішим і унікальним ресурсом суспільства, розглядається як міра розуму, інтелекту та цілеспрямованості дій, носій соціального середовища, вищої форми відображення матерії (свідомості); ***організація*** *(або організованість)* є формою ресурсів у соціумі, групі, яка визначає його структуру, включаючи інститути людського суспільства, його надбудови, є мірою впорядкованості ресурсів; організація системи пов'язана з наявністю деяких причинно-наслідкових зв'язків у цій системі і може мати різні форми, наприклад, біологічну, інформаційну, екологічну, економічну, соціальну тощо; ***простір*** – міра протяжності матерії, розподілу її в навколишньому середовищі; ***час*** – міра незворотності змін матерії (подій).

Можна говорити, що людина «перебуває» одночасно в різних полях – матеріальному, енергетичному, інформаційному, соціальному, проектуючи на себе їх просторово-часові та ресурсні (матерія, енергія, інформація) характеристики.

Приклад. Розглянемо просту задачу – піти вранці на заняття до університету. Це завдання, яке мало не щодня вирішує студент, має усі, згадані вище, аспекти:

- матеріальний – студенту необхідно перемістити деяку масу, наприклад підручники і зошити, на потрібну відстань;
- енергетичний – студенту необхідно мати запас і витратити певну кількість енергії на переміщення;
- інформаційний – необхідна інформація про маршрут руху і розташування університету, її потрібно обробляти в процесі свого руху;
- людський (соціальний) – переміщення, зокрема, пересування на автобусі, неможливе без водія автобуса;
- організаційний – необхідна наявність відповідних транспортних мереж та маршрутів, облаштованих зупинок тощо;
- просторовий – переміщення на деяку відстань;
- часовий – на це переміщення буде витрачено час (за який відбудуться деякі незворотні зміни в середовищі, в стосунках, в зв'язках).

Всі типи ресурсів тісно пов'язані і переплетені, вони неможливі один без одного, актуалізація одного з них веде до актуалізації інших. Якщо класичне природознавство пояснює світ виходячи з руху, взаємоперетворення речовини і енергії, то зараз реальний світ, об'єктивна реальність можуть бути пояснені лише з урахуванням супутніх системних, і особливо системно-інформаційних та синергетичних процесів.

Особливий тип мислення – системний, властивий аналітику, який хоче не тільки зрозуміти суть процесу чи явища, але і керувати ним.

Предметне мислення дозволяє цілеспрямовано (як правило, з метою *вивчення*) виявити і пізнати причинно-наслідкові зв'язки та закономірності в низці подій і явищ. В цьому полягає *методика і технологія* дослідження систем.

Системне мислення дозволяє цілеспрямовано (як правило, з метою *управління*) виявити і пізнати причинно-наслідкові зв'язки та закономірності в низці подій і явищ. В цьому полягає *методологія* дослідження систем.

При системному мисленні сукупність процесів (які можуть складатися з різних складових елементів) актуалізується, досліджується як єдине ціле, як одна організована за загальними правилами система, поведінку якої можна передбачити без з'ясування поведінки складових елементів, вивчення їх якості та кількості. Поки не стане зрозуміло, як функціонує чи розвивається система в цілому, ніякі знання про її частини не дадуть повної картини цього розвитку.

Приклад. Відповідно до принципу системного мислення суспільство складається з людей. Кожна людина – також система (фізіологічна, наприклад). У людини, в свою чергу, є властиві їй як організму системи, наприклад, система кровообігу. Коли люди взаємодіють з іншими людьми, утворюються нові системи – сім'я, етнос та ін. Ця взаємодія може відбуватися на рівні громадських інституцій чи окремих людей і навіть окремих систем кровообігу (наприклад, при прямому переливанні крові).

Відповідно до принципу системного підходу, кожна система впливає на іншу систему. Весь навколишній світ – це взаємодіючі системи. Мета системного аналізу – з'ясувати ці взаємодії, їх потенціал і використати їх для потреб людини.

Предметний аналітик – професіонал, який вивчає, описує деяку предметну область, проблему відповідно до принципів, методів та технологій цієї галузі. Це не означає «вузький» розгляд проблеми, хоча саме так часто трапляється.

Системний аналітик – професіонал високого рівня, що вивчає, описує системи відповідно до принципів системного підходу, тобто вивчає проблему комплексно. Йому притаманний особливий склад розуму, що базується на мультизнаннях, широкому кругозорі і досвіді, високий рівень інтуїції, уміння приймати доцільні ресурсозабезпечені рішення. Його основне завдання – допомогти предметному аналітику прийняти правильне (узгоджене з іншими системами, таке, що не «погіршує» їх) рішення при розв’язанні предметних проблем.

Системним у світі є все: практика та практичні дії, знання і процес пізнання, навколишнє середовище та зв’язки з ним (в ньому). Системний аналіз як методологія наукового пізнання структурує все це, дозволяючи досліджувати і виявляти інваріанти об’єктів, явищ і процесів різної природи, розглядаючи спільне та відмінне, складне і просте, ціле і частини.

Будь-яка людська інтелектуальна діяльність має бути за своєю суттю системною діяльністю, що передбачає використання сукупності взаємопов’язаних системних процедур на шляху від постановки задачі, цілей, планування ресурсів до знаходження і впровадження рішень.

Приклад. Будь-яке економічне рішення має базуватися на фундаментальних принципах системного аналізу, економіки, інформатики, управління та враховувати поведінку людей у відповідному соціально-економічному середовищі.

Невикористання системного аналізу не дозволяє знанням (що закладаються традиційною освітою) перетворюватися на вміння і навички їх застосування, в фахові компетентності з побудови і реалізації цілеспрямованих, структурованих, забезпечених ресурсами конструктивних процедур вирішення проблем. Системно мисляча і діюча людина, як правило, прогнозує і аналізує результати своєї діяльності, порівнює свої бажання (цілі) і свої можливості (ресурси), враховує інтереси навколишнього середовища, розвиває інтелект, виробляє правильний світогляд і правильну поведінку в людських колективах.

Світ, що оточує нас, нескінченний у просторі і в часі; людина ж розпоряджається скінченними ресурсами (матеріальними, енергетичними, інформаційними, людськими, організаційними, просторовими і часовими). Протиріччя між безмежністю бажання людини пізнати світ і обмеженою (ресурсами, невизначеністю) можливістю зробити це, між нескінченністю природи і скінченністю ресурсів людства, мають багато важливих наслідків, зокрема, і для самого процесу пізнання людиною навколишнього світу. Однією з таких особливостей пізнання, яка дозволяє поетапно вирішувати ці суперечності є використання аналітичного та синтетичного способу мислення, тобто поділу цілого на частини та уявлення складного як сукупності простіших компонентів, і навпаки, з'єднання простих елементів та побудови, таким чином, складного.

Приклад. Аналітичність людського знання проявляється в існуванні різних наук, в більш глибокому вивченні конкретних вузьких питань, кожне з яких є і цікавим, і важливим, і необхідним. Проте, таким же необхідним є і зворотний процес синтезу знань. Так виникають «прикордонні» науки – синергетика, біоніка, біохімія та ін. Ще більш висока форма синтетичних знань реалізується в науках про найзагальніші властивості природи: філософія виявляє і описує загальні властивості всіх форм матерії; математика також вивчає деякі загальні відносини. До синтетичних наук відносяться системний аналіз, інформатика, кібернетика та ін., що поєднують формальні, технічні, гуманітарні та інші знання.

Процес пізнання є системним, він структурує в системи оточуючий нас світ. Все, що не пізнане на даний момент, утворює «хаос в системі», який, будучи непізнаним в рамках даної теорії, змушує шукати нову інформацію, нові форми подання та опису знань, призводить до появи нових наукових напрямів, дає стимул для розвитку умінь та навичок дослідника.

При викладенні основ теорії систем можливі два основні підходи: формальний і понятійно-змістовний. *Формальний підхід* використовує різного рівня математичний апарат – від простих співвідношень до операторів, функціоналів, категорій, алгебр. *Понятійно-змістовний підхід* – базується на основних поняттях, ідеях, концепціях, методологічних принципах. Багато ідей і принципів системного аналізу, хоча й більш точні та строгі у викладі формальною мовою, проте, зберігають свою силу та актуальність і на змістовному рівні. Крім того, чинник невизначеності в системному аналізі обмежує застосовність строгих математичних формулювань.

1.2. Поняття системи

Під **системою** S розуміють множину взаємозв'язаних елементів будь-якої природи, які поєднані за деякими системоутворюючими ознаками та підпорядковані спільній меті.

Приклад. Університет як система з погляду ректора, головного бухгалтера, начальника служби охорони, студента складається з різних підсистем та елементів.

Незважаючи на інтуїтивну зрозумілість та велику важливість терміну «система» для наукових досліджень, донині не існує єдиного загальноприйнятого його визначення. Можна виділити різні змістові аспекти цього поняття:

- у «конструкторському» розумінні «система» подається як проектування та створення певного комплексу методів і засобів, які дослідник або розробник застосовує для досягнення певної мети, для виконання свого завдання;
- в науково-дослідницькому трактуванні «система» уявляється як загальна методологія дослідження процесів і явищ, що відносяться до певної галузі людських знань;
- у теоретико-пізнавальному аспекті «система» розуміється як спосіб мислення.

Математична модель є спрощеним абстрактним аналогом системи, вираженим за допомогою формул та рівнянь.

Зовнішнє середовище E даної системи складають елементи, які не належать цій системі, але які впливають на систему або змінюються під її впливом.

Приклад. Весь наш світ можна розглядати як гігантську систему, але ми не досліджуємо Всесвіт практично кожен раз, коли виникає проблема. Тому певна система є підсистемою Всесвіту, а Всесвіт лише в найбільш широкому сенсі можна називати середовищем цієї системи, а в абсолютній більшості випадків середовище – це все те, що взаємодіє з системою, тобто теж певна підсистема Всесвіту.

Система взаємодіє із зовнішнім середовищем за допомогою своїх «входів» та «виходів».

Вхід системи – це канал, за допомогою якого зовнішнє середовище E впливає на систему S. Через входи до системи надходить речовина, енергія, інформація, фінанси.

Вихід системи – це канал впливу системи S на зовнішнє середовище. Речовина, енергія чи інформація, зазнавши деяких перетворень, надходять до зовнішнього середовища через «вихід».

Загальна кількість взаємодій системи з зовнішнім середовищем дуже велика, тому на практиці обмежуються аналізом найсуттєвіших зв'язків, вибір яких визначається конкретними умовами управління тим чи іншим об'єктом. Позначивши множину входів $X = \{X_i\} = \{X_1, X_2, ..., X_n\}$, виходів – $Y = \{Y_j\} = \{Y_1, Y_2, ..., Y_m\}$, а перетворюючі відношення між ними – R, запишемо систему у символьному вигляді: YRX. Тобто, систему можна розглядати як перетворювач входів на виходи: $Y = RX$ (рис. 1.1).

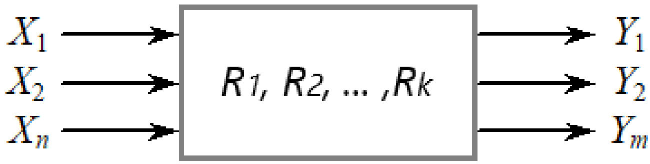


Рис. 1.1. Схема системи «вхід-вихід»

Приклад. Розглянемо банк як систему. Зовнішнє середовище банку – це система інвестицій, фінансування, трудових ресурсів, нормативів тощо. Входи – характеристики (параметри) цієї системи. Стан системи – показники фінансового стану банку. Виходи – потоки кредитів, послуг, виплат тощо. Функції системи – це банківські операції, наприклад, кредитування.

Підсистемою називають сукупність елементів, які об'єднані єдиним процесом функціонування та при взаємодії реалізують певну операцію, що необхідна для досягнення поставленої перед системою в цілому мети.

Надсистемою називають ширшу систему, в яку входить досліджувана система як складова частина.

Елементом системи називають її частину, яка виконує специфічну функцію і є неподільною з погляду завдання, що розв'язується. Внутрішня структура елементів не є предметом системного аналізу. Важливі лише властивості елемента, які визначаються його взаємодією з іншими елементами системи та справляють вплив на поведінку системи.

Слід зауважити, що поділ системи на елементи та саме поняття елемента є певною мірою відносним і умовним.

Між елементами довільної системи та між різними системами існують **зв'язки**, за допомогою яких вони взаємодіють між собою. Ці зв'язки можуть виражатися обміном речовиною, енергією чи інформацією між взаємодіючими системами або елементами. Система може мати зовнішні та внутрішні зв'язки. Зв'язки можуть бути також як прямими, так і зворотними.

Системи мають зовсім нові якості, які відсутні у її елементів. Ці якості виникають саме завдяки наявності зв'язків між елементами. Саме за допомогою зв'язків здійснюється перенесення властивостей кожного елемента системи до інших елементів.

Зворотні зв'язки є складним механізмом причинної залежності та полягають у тому, що результат попередньої дії впливає на наступний перебіг процесів у системі: причина підпадає під вплив зворотного впливу наслідку. Якщо зворотний зв'язок підсилює результат впливу наслідку, то його називають позитивним, а якщо послаблює – негативним. Негативні зворотні зв'язки сприяють збереженню стійкості системи. Тільки завдяки наявності зворотних зв'язків у системах можуть відбуватися процеси цілеспрямованої діяльності та регулювання.

Зв'язки перетворюють систему з простого набору компонентів у єдине ціле і разом з компонентами визначають стан та структуру системи, безумовно при визначальному впливі функції.

Важливим для опису систем є поняття структури. Під **структурою системи** розуміють її стійку впорядкованість та зв'язки між елементами і підсистемами. *Структура відбиває найсуттєвіші зв'язки між елементами та підсистемами, які мало змінюються при змінах у системі та забезпечують існування системи і найважливіших її властивостей.* Для визначення структури системи необхідно провести її послідовну ***декомпозицію***, тобто виділити в ній підсистеми всіх рівнів, доступних аналізу, та їх елементи, які відповідно до завдань дослідження не поділяються на складові частини. Завдяки ієрархічності структура складних систем може бути подана через структуру їх частин – від підсистем до елементів. Власне, суттю декомпозиції є спрощення системи, надміру складної для розгляду цілком.

Найпростіший спосіб формалізації структури – це відображення її графічно у вигляді графа, коли елементам системи відповідають вершини графа, а зв'язкам між елементами – дуги або ребра. Матричний спосіб передбачає подання структури за допомогою матриці суміжності чи інцидентності.

За ***топологією*** внутрішніх зв'язків структури розділяють на: послідовні, паралельні, радіальні, кільцеподібні, типу повний граф, деревоподібні, незв'язані тощо.

Приклад. Лінійну структуру можна уявити як послідовність автобусних зупинок на одному маршруті. Прикладом ієрархічної структури може бути система керування університетом «Ректор ~ Проректори ~ Директори ННІ ~ Завідувачі кафедр ~ Викладачі». Як приклад мережевої структури можна розглянути організацію робіт на будівництві: деякі роботи, наприклад, монтаж стін і благоустрій території можна виконувати паралельно. Приклад матричної структури –

система науковців відділу НДР, що працюють за однією тематикою.

Метою топологічного аналізу структури системи є відображення можливостей структури для реалізації функцій, виходячи з наявних елементів та відношень між ними, не вникаючи у їх змістовний опис. Отже, у загальному вигляді систему можна зобразити графічно у такий спосіб (рис. 1.2).

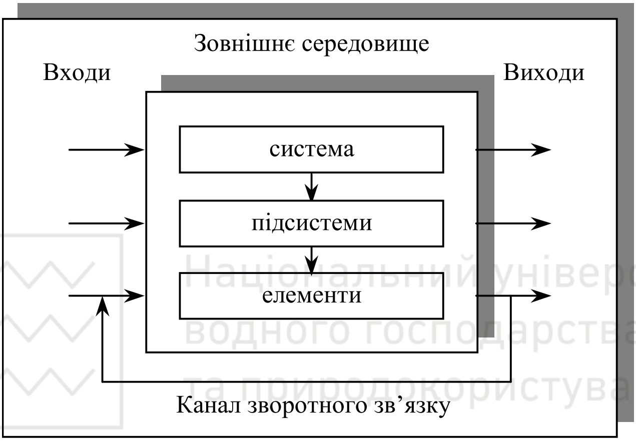


Рис. 1.2. Графічне зображення структури системи [30, С. 18]

За ***управлінням*** структури класифікують на: централізовані, децентралізовані, централізовані-розподілені, ієрархічні.

Централізована структура передбачає реалізацію усіх процесів керування об’єктами в одному органі керування, який безпосередньо отримує інформацію про стани об’єктів і передає керуючі сигнали кожному з них. Важливою перевагою такого способу керування є можливість організації глобально-оптимального управління. Недоліком цього методу є необхідність великих об’ємів засобів накопичення, високої продуктивності органів керування.

Децентралізована структура будується за умови незалежності об’єктів керування. Система з такою структурою складається з відносно незалежних між собою підсистем.

Недоліком такої структури є неможливість організації глобально-оптимального управління.

Особливістю ***централізованої-розподіленої*** структури є те, що в ній зберігається основна перевага централізованого керування, а саме – передача керуючих сигналів на основі аналізу інформації про стани всіх об'єктів. Але на відміну від централізованої структури орган керування є розподіленим. Недоліком цієї структури є складність інформаційної взаємодії між елементами розподіленого органу керування, а також складність синхронізації елементів органів управління.

Проблема організації інформаційної взаємодії розподіленого органу управління розв'язується у ***ієрархічній*** структурі. Ієрархічна структура – це структура з нерівноправними зв'язками, тобто з підпорядкованістю. Основним недоліком цієї структури є її «консервативність» (будь-яка зміна в структурі вимагає великої кількості зусиль).

Головним системоутворювальним чинником є функція. Під **функцією** системи будемо розуміти перетворення її входів у виходи. З іншого погляду, функція системи полягає у збереженні її існування, підтримці її структури та впорядкованості. Також функцією системи можна вважати функціонування системи, що визначається як спосіб, засіб або як сукупність дій, необхідних для досягнення мети.

Окрім функції система може мати мету (ціль). Загалом поняття «мета» відображає те, що може чи повинно виникнути, прообраз майбутнього, стан, який бажано досягнути. Пізнання мети допомагає зрозуміти сутність досліджуваної системи. **Мета системи** – це бажаний стан її виходів. Системи, що мають мету (ціль), називають **цілеспрямованими**. Будь-які соціально-економічні системи є цілеспрямованими, бо їх елементами є люди.

Мета має декілька аспектів: ***пізнавальний*** – відповідає прогнозу майбутнього, ***конструктивний*** – можливі способи переходу до бажаного майбутнього чи план дій.

Мета конкретизується шляхом побудови ***дерева цілей***. Цілі в *часовому* аспекті поділяються на тактичні цілі, макроцілі, та ідеали. ***Тактичні цілі*** – це бажані результати, досягнення яких відбувається за визначений і порівняно короткий період. ***Макроцілі*** досягаються за довший час і вимагають для цього досягнення кількох тактичних цілей. ***Ідеали*** – це такі цілі, які ніколи не досягаються, але до яких система постійно наближається, реалізуючи тактичні та макроцілі.

За *способом досягнення* цілі поділяються на функціональні, цілі-аналоги та цілі розвитку. ***Функціональна*** ціль – це ціль, спосіб досягнення якої відомий системі, що вже досягала цю ціль. Функціональні цілі повторюються в часі та просторі. Прикладами таких цілей є виконання виробничих операцій, що періодично повторюються, стандартні функції управління та ін. ***Ціль-аналог*** – це образ, отриманий в результаті дії іншої системи, але який ні разу не досягався системою, що розглядається, а якщо і досягався, то за інших умов. ***Ціль розвитку*** або нова ціль – це ціль, яка ніколи і ніким раніше не досягалася. Така ціль, по суті, пов'язана з утворенням нових систем.

Ці типи цілей пов'язані між собою. Ціль розвитку за умови її успішного досягнення однією з систем перетворюється в ціль-аналог для всіх інших систем, а для даної системи стає функціональною за умови незмінності зовнішніх умов чи ціллю-аналогом, якщо умови зміняться.

Зв'язок, що існує в системі чи поза нею, можна розглядати як деякий **потік**. Функції системи реалізуються через потоки енергії, людей, матеріальні та інформаційні. Структуру можна розглядати також як множину обмежень на потоки в просторі та часі. Структура ініціює потоки, спрямовуючи їх вздовж певних шляхів (каналів) та перетворює їх з деякою затримкою в часі (час перетворення).

Найважливішими потоками у складних системах є ***інформаційні***, оскільки вони супроводжують усі інші потоки і, окрім цього, часто є домінуючими. Відображення структури системи, у якій зв'язками є інформаційні потоки, можна здійснити за допомогою ***діаграм потоків даних*** DFD (data flow diagrams), які були запропоновані De Marco (1978 р.) і доповнені Gane та Sarson (1979 р) [9, С. 28–32]. Ці діаграми використовують для аналізу та моделювання інформаційних систем з метою мінімізації потоків даних та зменшення їх обсягу, уникнення як дублювання інформації, так і дублювання шляхів її передачі. DFD відображають джерела та споживачів інформації, вид і напрямок передачі даних, елементи накопичення та процеси перетворення.

Стан системи – це зафіксовані значення характеристик системи, важливих для цілей дослідження. Стан системи визначається *кількісними* та/або *якісними* значеннями внутрішніх параметрів (змінних) системи в даний момент. Зміна будь-якої з цих характеристик означає перехід системи до іншого стану. Функціонування системи або зміну станів системи у часі називають **поведінкою** або **рухом**. Таким чином, поведінка системи – це розгорнута в часі послідовність реакцій системи на внутрішні зміни та зовнішній вплив. **Процес** – це набір (послідовність) станів системи, що відповідає впорядкованій *неперервній* або *дискретній* зміні деякого параметра, який визначає характеристику чи властивість системи. В більшості випадків таким параметром є час. Процес зміни станів системи в часі відображає її **динаміку**.

Важливими властивостями системи є комунікативність, інтегративність, ступінь рівноваги та стійкості, адаптація. ***Комунікативність*** системи – це ступінь зв'язку із зовнішнім середовищем. Поняття ***інтегративності*** визначає чинники, які утворюють і зберігають систему. ***Рівновага*** – це здатність системи зберігати свій стан якомога довше (як за відсутності, так і за наявності зовнішніх збурюючих впливів). Під ***стійкістю*** розуміють здатність системи повертатися в стан рівноваги після виведення її з цього стану впливом зовнішніх збурень. Стан рівноваги, у який система здатна повертатися, називають *стійким станом рівноваги*. ***Адаптація*** – здатність системи до цілеспрямованого пристосування.

1.3. Класифікація систем. Складні системи

В залежності від мети дослідження можна обрати різні принципи та підходи до класифікації систем (рис. 1.3). Загалом, системи можна розділити на ***абстрактні*** та ***матеріальні***. Елементами матеріальних систем є реальні предмети і явища;

елементи абстрактних систем – це символи (знаки, букви, цифри). Матеріальні системи за походженням поділяють на ***природні*** та ***штучні***. Природні (біологічні, фізичні, хімічні, екологічні тощо) системи є сукупністю природних об'єктів (ліс, атом, молекула, організм, популяція). Природні системи поділяють на ***живі*** та ***неживі***. Створені людиною штучні системи включають як ***формальні*** системи (мови, математичні моделі), так і ***неформальні***, які поділяють на ***технічні*** та ***системи за участю людини*** (людино-машинні, соціально-економічні тощо).

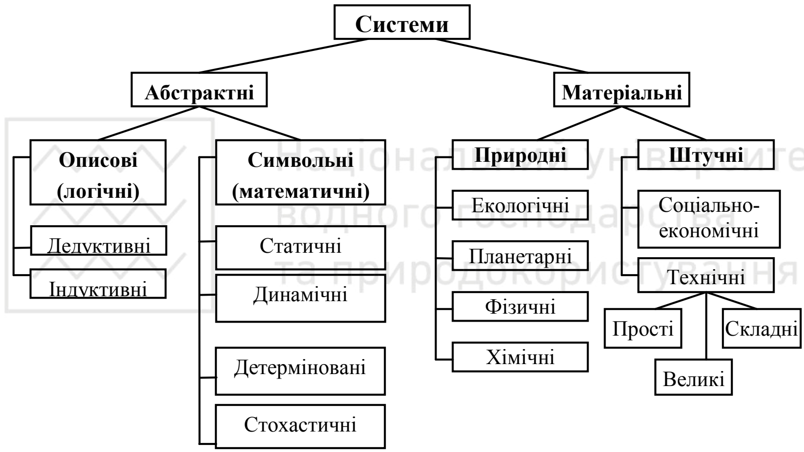


Рис. 1.3. Класифікація систем

За призначенням системи поділяються на ***пасивні*** (пристрої, які мають деяке функціональне призначення, наприклад, автомобіль – для транспортування, будинок – для захисту від зовнішнього середовища) та ***активні*** (цілеспрямовані). ***Цілеспрямовані*** системи спроможні самі визначати свою поведінку завдяки зовнішньому керуванню або внутрішньому «прагненню» системи до деякої цілі.

З точки зору взаємодії із зовнішнім середовищем системи поділяють на ***відкриті*** та ***ізольовані*** (***автономні***). Ізольовані системи майже не взаємодіють із зовнішнім середовищем і з часом досягають стану рівноваги. Цілеспрямовані системи завжди є відкритими, оскільки обмінюються матерією, енергією або інформацією з оточуючим середовищем.

Розрізняють ***статичні*** та ***динамічні*** системи. У статичній системі фіксуються статичні взаємовідношення на певний момент. Системи статичної структури корисні для створення теоретичної бази з метою подальшого аналізу та синтезу систем. Якщо система переходить із часом від одного стану до іншого, то такі системи називають динамічними.

Системи поділяються також на ***детерміновані*** та ***стохастичні***. У детермінованих системах перехід з одного стану в інший (тобто поведінка системи) є визначеним. На відміну від детермінованих систем рух (розвиток) стохастичних систем не є чітко визначеним та розглядається як випадковий процес.

За типом та характером взаємодії між елементами системи і системою та зовнішнім середовищем розрізняють ***лінійні*** і ***нелінійні***, ***ієрархічні***, ***керовані*** системи, системи що ***самоорганізуються*** тощо.

Приклад. Система «Комп'ютер» – штучна (технічна), відкрита, детермінована, керована ззовні (програмно) система з кількісно-якісними характеристиками (кількісні – швидкодія, розрядність, обсяг вінчестера чи оперативної пам'яті; якісні – дизайн, колір тощо). Система «Логічний диск» – абстрактна, відкрита, з кількісними характеристиками і змішаним управлінням. Система «Фірма» – штучна (соціально-економічна), відкрита, самокерована система, що адаптується. Екологічна система «Озеро» – природна, відкрита система з кількісно-якісними характеристиками (кількісні – ширина, глубина, хімічний склад води, координати розташування тощо; якісні – красота природи, неповторний краєвид).

Окремо слід виділити ***соціально-економічні*** системи – комплексні структури, що складаються із економічних, виробничо-технічних та соціальних підсистем, які виконують різні цілі (наприклад, місто, територіальна громада, організація, ОСББ, підприємство).

Одним із найбільш поширених класів систем є ***комп'ютеризовані (автоматизовані) системи управління***. За принципом функціонування такі системи поділяються на ***транзакційні*** (виконують прості операції перетворення між елементами вхідної інформації з метою формування вихідної), ***системи прийняття рішень*** (призначені для розв'язування складних задач, пов'язаних із перетворенням вхідної інформації у вихідну на основі певних правил (знань); як правило, мають елементи штучного інтелекту та вирізняються великою кількістю взаємопов'язаних алгоритмів функціонування) та ***геоінформаційні*** системи, у яких опрацювання просторової інформації здійснюється за допомогою графічного інтерфейсу на основі географічних, топографічних карт, планів.

Важливою класифікаційною ознакою систем є їх ***складність***. **Простою** можна вважати систему з невеликою кількістю взаємодіючих елементів, яка не має розгалуженої структури, а обсяг інформації, що необхідний для її описування та управління, є порівняно невеликим. Під **складною** розуміють систему, стан якої неможливо достатньо вичерпно та точно описати, вона має розгалужену структуру і значну кількість взаємодіючих елементів. Складну систему неможливо скомпонувати з окремих підсистем. Її можна зрозуміти лише аналізуючи різними методами – фізичними, хімічними, математичними тощо.

Приклади: людський мозок, хімічні реакції на молекулярному рівні, метаболізм біологічних клітин, економічна система країни, людське суспільство.

Розрізняють також **великі** системи – системи, моделювання яких ускладнено через їх велику розмірність. Як правило, дослідження таких систем можливе лише як сукупності підсистем.

Існують різні підходи до оцінки **складності систем**:

- ***алгоритмічна*** концепція, що визначає складність як довжину алгоритму відтворення системи;

• ***обчислювальна*** концепція, що пов'язує алгоритмічну складність та обчислювальні ресурси;

• ***теоретико-множинна*** концепція, що ототожнює складність системи з числом її елементів;

• ***теоретико-інформаційна*** концепція, що пов'язує складність системи з її ентропією.

Алгоритмічна складність задає складність описання алгоритму розв'язання задачі. Алгоритмічна складність доповнюється обчислювальною, яка характеризує витрати різних обчислювальних ресурсів на розв'язування заданого класу задач. Міру обчислювальної складності характеризують також: надійність обчислень; можливість розпаралелювання обчислювального процесу; частота звертань до пристроїв комп'ютера та розподіл даних між зовнішньою і оперативною пам'яттю.

У рамках теоретико-інформаційної концепції Вільям Росс Ешбі запропонував оцінювати складність через різноманітність системи, що кількісно оцінюється кількістю можливих станів системи n

$$H_m = log_2 n . \quad (1.1)$$

Оцінка складності В. Р. Ешбі за кількістю станів не враховує ймовірність, з якою система може перебувати в даному стані. Нехай P_i є ймовірність того, що система знаходиться в i-тому стані. Тоді для оцінки складності системи можна використовувати формулу Клода Шеннона

$$H = -\sum_{i=1}^{n} P_i \cdot log_2 P_i . \quad (1.2)$$

Властивості складних систем

Властивість ***взаємної автономності*** елементів системи проявляється в тому, що кожному її елементу притаманні властивості системи в цілому.

Іманентність виражається в тому, що системоутворююче відношення властиве лише для елементів даної системи.

Множинність – одна і та ж сукупність елементів може бути множиною різних систем, які відрізняються системо-утворюючими властивостями та конкретними відношеннями між елементами. (*Приклад*. Студентська група може розглядатися як навчальна система в освітньому процесі університету і як соціальна система взаємостосунків у колективі).

Надійність системи – це здатність зберігати системо-утворюючу властивість при елімінації (вилученні) деякої кількості елементів.

Завершеність системи виявляється в тому, що вона не допускає приєднання нових елементів без руйнування цієї системи.

Мінімальність. Якщо система руйнується при вилученні хоча б одного елемента, то система мінімальна.

Цілісність – система поводить себе як єдине ціле. Це означає, що, з одного боку, система – це цілісне утворення, а з іншого – в її складі чітко можуть бути виділені окремі об’єкти (елементи). Але не компоненти утворюють ціле (систему), а навпаки, при поділі цілого виявляють компоненти системи.

Первинність цілого – головний постулат теорії систем.

Емерджентність. Складна система має такі властивості, які не притаманні жодному з її елементів.

Синергізм. Ефективність спільного функціонування елементів системи є вищою, ніж сумарна ефективність ізольованого функціонування цих же елементів.

Еквіпотенційність. Кожну систему можна розглядати як підсистему іншої більш крупної системи; кожен елемент системи, в свою чергу, є системою.

До властивостей, що формують ***здатність системи до самозбереження*** належать:

- здатність зберігати ***рівноважний стан*** незалежно від умов зовнішнього середовища;
- здатність самостійно утримувати основні параметри в ***допустимих межах*** (цю властивість називають ***гомеостазом***);
- здатність динамічно реагувати на зміни й впливи навколишнього середовища за рахунок ***структурних перебудов***.

Відкриті складні системи мають низку характерних особливостей:

- ***диференціація*** – різні складові системи виконують різні функції і не є взаємозамінними. *Наприклад*: економічна система – сторож, бухгалтер, програміст;
- ***централізація*** **–** з часом одна із складових системи може стати домінуючою (*приклад*: урядова бюрократія вимагає концентрації та централізації влади; для корпоративного стилю управління більш характерною є децентралізація владних повноважень);
- ***цілеспрямованість*** – процеси можуть протікати по різному, але кінцевий стан системи буде однаковим (*приклад*: струмки, що стікають зі схилу гори в озеро, яке розташоване в низині), тобто, поведінка системи та її структура спрямовані на досягнення певної мети;
- ***історичність*** **–** система не може бути незмінною, вона виникає, функціонує, розвивається і гине.

1.4. Синергетика. Принципи синергетики

Системи, що ***самоорганізуються*** – це системи, здатні протистояти ентропійним тенденціям, здатні адаптуватися до зовнішніх умов, перетворюючи свою структуру.

Якщо в основі системного аналізу лежить принцип системності, то в основі теорій самоорганізації – принцип розвитку. Обидва ці принципи взаємодоповнюють один одного і насправді утворюють єдність, що відображається в пізнанні як єдність теорій самоорганізації та системних досліджень. До теорій самоорганізації належать синергетика, теорія змін і теорія катастроф.

Синергетика, основні положення якої були сформульовані професором Штутгартського університету Г. Хакеном, є евристичним методом дослідження відкритих систем, що самоорганізовуються, схильних до кооперативного ефекту, який супроводжується утворенням просторових, часових або функціональних структур; або, стисло, процесів самоорганізації систем різної природи. Вона виникла у відповідь на кризу стереотипного, лінійного мислення, що вичерпало себе.

Терміном «**синергетика**» – (від грец. *synergetikos* – спільний, погоджений, діючий) називають науковий напрямок, який вивчає зв'язки між елементами структури (підсистемами), які утворюються у відкритих складних системах (біологічних, фізико-хімічних, економічних та ін.) завдяки інтенсивному (потоковому) обміну речовинами і енергією з навколишнім середовищем за нерівноважних умов. Теоретичні засади синергетики – термодинаміка нерівноважних процесів, теорія випадкових процесів, теорія нелінійних коливань і хвиль тощо.

Отже, **синергетика** – це міждисциплінарний науковий напрямок, що вивчає закономірності процесів самоорганізації, еволюції та кооперації. ***Мета синергетики*** полягає в побудові загальної теорії складних систем.

Як зазначалося вище, дослідження складних систем суттєво ускладнене через низку особливостей таких систем, які, на відміну від простих, мають наступні характеристики:

- відкритість системи;
- активність (цілеспрямованість) компонентів;
- багато паралельних взаємозв'язків між елементами;
- кооперативна поведінка компонентів;
- здатність до навчання;
- невизначеність параметрів середовища.

Розвиток складних системи – це поетапне проходження ***точок біфуркації*** (***роздвоєння***). Поблизу точок біфуркації спостерігається різке посилення ***флуктуацій*** (***коливань***). Зона біфуркації характеризується принциповою непередбачуваністю – невідомо, чи стане подальший розвиток системи хаотичним, чи народиться нова, більш впорядкована структура. Можливість спонтанного виникнення порядку з хаосу – найважливіший момент процесу самоорганізації у складній системі. Одним із напрямків синергетики є математичне моделювання переходу систем з одного стійкого стану в інший.

Синергетичний підхід визнає можливість різноманітного розвитку систем, наявність біфуркації. У точці біфуркації нестійкість підсилюється через те, що в системах завжди присутні флуктуації, які гасяться у стійкому стані. Але в результаті нелінійних процесів, які виводять параметри системи за критичні значення, такі флуктуації посилюються і можуть спричинити стрибкоподібний перехід до нового стійкого стану з меншою ентропією, після чого цикл «плавний розвиток – стрибок», «еволюція – революція», «стійкість – нестійкість» повторюється.

Таким чином, стійкість та нестійкість є однаково необхідними у процесі розвитку будь-якої системи. Абсолютно хистка система не може протистояти флуктуаціям, не здатна до адаптації і швидко руйнується. Проте надто стійка система, придушуючи будь-які флуктуації, консервує свою структуру і поводження, а тому не здатна змінитися якісно, вона позбавлена можливості розвитку, і її руйнування стає лише справою часу. Обидва типи систем приходять до хаосу, різниця між ними полягає лише в часі, що проходить до вибухового зростання ентропії.

Загалом, синергетикою називають сукупність знань про хаос і порядок, перехідні процеси, фрактали і нелінійність, які розуміють і як теорію, і як навчання, і як науку, і як світогляд, виходячи з різних образів, фактів, уявлень про хаос, порядок, когерентність, перехідних і кооперативних процесів у природі, суспільстві, духовному світі. Перелік ідей, що формують синергетику як парадигму, містить нелінійність, самоорганізацію, відкритість системи, її нерівноважність тощо [6, С. 91–95].

Синергетика як теорія самоорганізації виходить з того, що складним системам не можна нав'язувати шляхи їх розвитку, а швидше необхідно зрозуміти, як сприяти їх власним тенденціям розвитку, як виводити системи на ці шляхи, зрозуміти закони сумісного життя природи і людства, їх коеволюції. Для складних систем існує декілька альтернативних шляхів розвитку, вибір яких залежить від результату боротьби протилежних сил. Актуальним є наукове обґрунтування цього шляху.

З погляду синергетики неефективне управління полягає в нав'язуванні системі такої поведінки, яка їй не властива. Згідно із синергетичною концепцією більш ефективним буде так зване «м'яке» управління (на відміну від «жорсткого», програмного). *М'яке управління* – це управління за допомогою незначних, але належних резонансних впливів, які мають відповідати власним внутрішнім тенденціям розвитку системи. Головна мета такого управління полягає в тому, щоб завдяки незначному резонансному впливу «підштовхнути систему» до одного із її власних сприятливих шляхів розвитку. Своєчасні резонансні впливи можуть виявити значні, потужні внутрішні резерви системи [6; 12].

Головні принципи синергетики

Принципи ***синергетичної методології*** можна розбити на три групи: ***принципи складності*** (1–3), ***принципи невизначеності*** (3–6) і ***принципи еволюції*** (7–9).

Нільс Бор

1. Принцип додатковості Н. Бора. У складних системах виникає необхідність поєднання раніше несумісних, а згодом взаємодоповнюючих моделей і методів опису (квантово-хвильовий дуалізм).

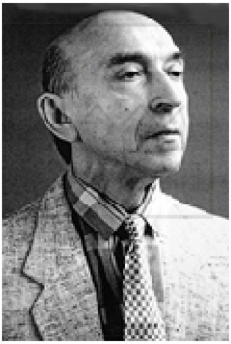
2. Принцип спонтанного виникнення І. Р. Пригожина. У складних системах можливі критичні стани, коли найменші відхилення можуть призвести до появи нових структур, повністю відмінних від попередніх (це може вести до катастрофи – ефект «снігової кулі»).

3. Принцип несумісності Л. Заде. При зростанні складності системи зменшується можливість її точного описання. Точність і змістовність інформації стають несумісними характеристиками.

4. Принцип незнання. Знання про складні системи принципово є неповними, неточними і суперечливими: Вони формуються на основі індивідуальних думок та колективних ідей.

Ілля Романович Пригожин

5. Принцип різноманітності шляхів розвитку. Розвиток складної системи є багатоваріантним. Критичний момент майбутнього розвитку складної системи пов'язаний з наявністю зон біфуркації – «розгалуження» можливих шляхів еволюції системи.



Лотфі Заде

6. Принцип пульсуючої еволюції. Процес еволюції складної системи носить не поступальний, а циклічний характер: він поєднує етапи зростання різноманітності та етапи згортання різноманітності, фази зародження порядку і підтримки порядку. Відкриті складні системи пульсують: розбігання змінюється зближенням, послаблення зв'язків – їх посиленням тощо.

7. Закон простоти складних систем. Реалізується і виживає той варіант складної системи, який має найменшу складність. Закон простоти складних систем реалізується природою в низці конструктивних принципів: ієрархічної побудови складних систем; симетрії; польової взаємодії (взаємодії через носій поля).

Курт Гедель

8. Теорема К. Геделя про неповноту. У багатьох теоріях (включаючи арифметику) завжди існують істинні твердження, які неможливо довести логічно, опираючись на систему аксіом.

9. Закон еквівалентності варіантів побудови складних систем. З ростом складності системи частка варіантів її побудови, близьких до оптимального, зростає.

1.5. Кібернетичні системи

З кібернетикою пов'язаний розвиток таких системних уявлень, як типізація моделей систем, виявлення особливого значення зворотних зв'язків у системі, підкреслення принципу оптимальності в управлінні та синтезі систем, усвідомлення значення інформації та можливостей її кількісного описання, розвиток методології моделювання, особливо проведення обчислювальних експериментів із застосуванням комп'ютера (що привело до розвитку важливого напрямку моделювання — імітаційного). Взаємовплив кібернетики і системного аналізу породив новий великий клас систем – кібернетичні системи, характерною особливістю яких є ***управління***.

Управління системою необхідне для забезпечення її цілеспрямованої поведінки при зміні умов зовнішнього середовища або умов її функціонування. Управління досягається за рахунок відповідної організації системи, під якою розуміють її структуру та спосіб функціонування. Системи з управлінням називають ***кібернетичними системами****.*

Взаємодія системи з зовнішнім середовищем свідчить, що середовище надає системі ресурси, а одержує від неї та споживає продукти кінцевої діяльності системи. Ці кінцеві продукти не можуть бути створені в середовищі (принаймні в достатній кількості), оскільки за таких умов немає необхідності виділяти систему із середовища. Система необхідна середовищу для задоволення деяких своїх потреб в її кінцевих продуктах. Тому можна зробити висновок, що до створення нових систем спонукає наявність незадоволених потреб, або, інакше кажучи, система створюється для вирішення деякої проблемної ситуації.

Отже, об'єктивною основою формування системи є проблемна ситуація, тобто такий незадовільний стан елементів зовнішнього середовища, який середовище власними засобами (сукупністю систем зовнішнього середовища) на даному етапі не в змозі нормалізувати.

Кібернетичні системи – це складні динамічні системи з управлінням. Інформація, яка поступає у систему, є засобом впливу на поведінку системи. За способом управління системи поділяють на ***керовані ззовні***, ***самокеровані*** та із ***комбінованим управлінням***.

Управління – це цілеспрямоване втручання в перебіг процесів у системі з метою деяких змін або ж підтримання стану системи незмінним. В технічних системах управляючими сигналами виступають потоки речовини та енергії. В економіці інструментами управління є потоки речовини, фінансів та інформації. В соціальних системах управління здійснюють за допомогою потоків інформації (накази, поради, стимули) та змін структури (організація нових груп і колективів).

Елементи ***кібернетичної системи*** здатні сприймати, запам'ятовувати і переробляти інформацію, а також обмінюватися нею. *Приклади кібернетичних систем*: автопілот, комп'ютер, людський мозок, живий організм, підприємство, суспільство.

Для складних кібернетичних систем притаманні такі властивості як лінійність та нелінійність, детермінованість та стохастичність, стаціонарність та ергодичність.

Лінійність чи ***нелінійність*** системи визначається її статичною характеристикою. Під статичною характеристикою системи розуміють зв'язок між величиною зовнішнього впливу $x(t)$ на систему (величиною вхідного сигналу) і максимальною величиною (амплітудою) вихідної характеристики y_m. Якщо функція $y_m = f(x)$ лінійна, то і система лінійна. Поняття «лінійності» означає наявність пропорційності між вхідними та вихідними параметрами. *Приклад лінійної функції*: зв'язок між температурними шкалами Цельсія (t) і Фаренгейта (T)

$$T = {}^{9}\!/_{5}\, t + 32\,. \qquad (1.3)$$

Нелінійність характеристик і наявність запізнювання в реагуванні є ознакою нелінійності системи. В математичному сенсі «нелінійність» означає певний вид рівнянь, що містять шукані величини в степенях більших, ніж одиниця, або коефіцієнти, що залежать від властивостей середовища. Нелінійні рівняння можуть мати декілька якісно різних розв'язків. Множині розв'язків нелінійного рівняння відповідає множина шляхів еволюції системи, що описується рівняннями нелінійної системи. Більше того, вивчаючи різні стадії розвитку процесів у відкритому нелінійному середовищі, можна очікувати на якісну зміну процесів, зокрема, переструктурування – ускладнення чи деградацію – організації середовища. Причому це відбувається не через зміну параметрів середовища, а як результат розвитку процесів у ньому.

Як *приклад* нелінійної системи розглянемо систему Лоренца

$$\begin{aligned} X' &= -\sigma X + \sigma Y, \\ Y' &= -XZ + \lambda X - Y, \\ Z' &= XY - bZ, \end{aligned} \tag{1.4}$$

де *X(t)*, *Y(t)*, *Z(t)* – динамічні змінні, що описують відповідно інтенсивність руху, різницю температур висхідного та низхідного потоків рідини і відхилення вертикального розподілу температури від лінійного режиму; σ, λ, b – деякі параметри.

Ця система рівнянь отримана Едвардом Лоренцом у 1963 р. при моделюванні конвекційного руху в шарі рідини, що підігрівається знизу. Її розв'язок, названий ***ди́вним атра́ктором Ло́ренца***, демонструє хаотичну поведінку (рис. 9.6).

За обумовленістю дії розрізняють системи з ***детермінованою*** і ***випадковою*** (***стохастичною***) дією. У детермінованій системі її складові елементи і зв'язки між ними взаємодіють точно передбаченим способом. У випадковій системі її елементи взаємодіють таким чином, що не можна зробити точного передбачення її поведінки.

Детерміновані системи поділяють на *статичні* та *динамічні*. Для *статичної* системи середні арифметичні значення вихідного сигналу на різних відрізках часу не виходять за допустимі межі, обумовлені точністю вимірів. У *динамічній* системі середнє арифметичне значення вихідного сигналу на різних відрізках часу змінюється, оскільки в такій системі відбувається зміна стану її елементів з часом.

Випадкові (стохастичні) системи поділяються на *стаціонарні*, *нестаціонарні* та *ергодичні*. Такий розподіл систем заснований на різній залежності від часу основних статистичних характеристик.

Система є ***стаціонарною*** (у вузькому розумінні), якщо відсутня зміна в часі математичного сподівання і дисперсії вихідного сигналу. В деяких процесах стаціонарність зберігається протягом тривалого часу, в інших – лише на коротких відрізках часу. Розрізняють ергодичні і неергодичні стаціонарні системи. Для ***ергодичних*** систем усереднення за часом однієї реалізації вихідної характеристики на великому проміжку часу є рівносильним до усереднення великої кількості короткочасових реалізацій.

Теми для роздумів і контрольні запитання

1. Основні етапи історії розвитку системного підходу та теорії систем.
2. Що означають терміни гіберно, гібернет, кібернетика?
3. У чому сенс «дійсно ефективного керування» за Трентовським?
4. Системний підхід як поєднання логічного позитивізму, дедукції та індукції, редукціонізму, казуальної логіки.
5. Предметом системного аналізу є…
6. Труднощі сприйняття проблемних ситуацій людьми.
7. Системний аналіз та інші методи наукових досліджень.
8. Взаємозв’язок між кібернетикою та загальною теорією систем.
9. Системний підхід як синтез індуктивного та дедуктивного способів мислення.
10. Аналіз потреб та областей застосування системного аналізу.

11. Особливості раціонального сприйняття складних систем.
12. Наведіть приклади взаємозв’язку просторово-часових та ресурсних (матерія, енергія, інформація, суспільство) характеристик деякої проблемної ситуації.
13. Чим відрізняються предметне і системне мислення?
14. Поняття системи. Різні трактування терміну «система».
15. Схема системи типу «вхід-вихід». Наведіть приклади.
16. Поняття про зв’язки в системі. Зворотні зв’язки.
17. Структура системи. Елементи і підсистеми. Декомпозиція. Топологічний аналіз структури системи.
18. Системи з централізованим, децентралізованим, централізовано-розподіленим управлінням та ієрархічні.
19. Що таке функція і мета системи? Як класифікуються цілі?
20. Стан та поведінка системи, процес, динаміка системи.
21. Класифікація і властивості систем.
22. Прості, складні та великі системи. Оцінка складності.
23. Поняття про синергетику. Головні принципи синергетики. Наведіть приклади, що ілюструють принципи синергетики.
24. Що таке кібернетичні системи? Наведіть приклади.
25. Поняття управління системою.
26. Властивості складних кібернетичних систем.
27. Поясніть значення понять: лінійність та нелінійність, детермінованість та стохастичність, стаціонарність та ергодичність систем.

2. МЕТОДОЛОГІЯ СИСТЕМНИХ ДОСЛІДЖЕНЬ

2.1. Системний підхід та системний аналіз

*Загалом **головне завдання системних досліджень** полягає в пошуку ефективних методів та засобів дослідження і управління об'єктами.* Це означає: виявлення та чітке формулювання проблеми за умов невизначеності; визначення або вибір оптимальної структури системи; виявлення цілей функціонування та розвитку систем; вивчення організації взаємодії між підсистемами та елементами; врахування впливу зовнішнього середовища; вибір оптимальних алгоритмів функціонування системи.

Системний підхід враховує взаємозв'язок явищ навколишнього світу. Він вимагає розглядати досліджувані явища і об'єкти не тільки як самостійну систему, але і як підсистему деякої великої системи. Основними *аспектами* системного підходу є:

- елементний (з яких елементів утворена система);
- структурний (розкриває внутрішню структуру системи, спосіб взаємодії її елементів);
- функціональний (показує, які функції виконує система та її елементи);
- комунікаційний (розкриває взаємозв'язок даної системи з іншими як по горизонталі, так і по вертикалі);
- інтегративний (показує механізми збереження та розвитку системи);
- еволюційний (відповідає на запитання як виникла система, які етапи проходила в своєму розвитку).

Принципи системного підходу – це положення загального характеру, що є узагальненням досвіду дослідження людиною складних систем. Їх часто вважають ядром методології. Найважливішими базовими принципами, на які спирається загальна теорія систем та системний аналіз, є *принцип системності* та *принцип ізоморфізму*.

Принцип системності відображає загальність погляду на об'єкти, явища і процеси світу як на системи з усіма властивими їм закономірностями. Цей принцип обумовлює необхідність спільного розгляду системи як цілого і як сукупності елементів, дослідження будь-якої частини системи разом з її зв'язками з іншими частинами та із зовнішнім середовищем.

Цей принцип постулює необхідність ієрархічного, принаймні трирівневого, дослідження системи: необхідно досліджувати власне систему, її підсистеми та елементи, а також розглядати систему як елемент системи вищого порядку.

Принцип ізоморфізму визначає наявність однозначної (*ізоморфізм*) чи часткової (*гомоморфізм*) відповідності структури однієї системи до структури іншої, що дає змогу моделювати одну систему за допомогою іншої, подібної в деякому відношенні. Сучасні дослідження як у загальній теорії систем, так і в тих галузях знань, які виникли на її основі (синергетика, теорія катастроф), свідчать про наявність не тільки ізоморфізму чи суворої відповідності структури систем, а й загального у їх розвитку та функціонуванні.

Обидва ці принципи підкреслюють наявність загальних системних закономірностей, але вони не виключають специфіки будови, функціонування та руху систем різних типів. Саме загальні закономірності намагається розкрити загальна теорія систем, тоді як аналізом загального й особливого в конкретних системах займаються інші галузі науки.

Серед інших важливих *принципів системного підходу* слід відмітити такі:

- ***принцип кінцевої мети***: глобальна мета системи має абсолютний пріоритет; будь-які зміни, удосконалення та акти управління повинні оцінюватися виходячи з того, чи сприяють вони досягненню кінцевої мети;
- ***принцип ієрархії***: корисним є введення ієрархії елементів та (чи) їхнє ранжування, виділення модулів (підсистем) у системі та розгляд системи як сукупності підсистем;
- ***принцип функціональності***: спільний розгляд структури і функцій системи з пріоритетом функцій над структурою; тобто, досліджувати структуру необхідно після зрозуміння функцій системи; при наданні системі нових функцій доцільно переглядати її структуру, а не прагнути «втиснути» нову функцію в стару структуру;

- ***принцип розвитку***: врахування динамічності системи, її здатності до розвитку, розширення, накопичення інформації, врахування невизначеності та випадковості при функціонуванні системи;
- ***принцип децентралізації***: співвідношення між централізацією та децентралізацією в управлінні системою; розумний компроміс між повною централізацією та наданням здатності реагувати на певні дії частинам системи.

Система з повною централізацією буде негнучкою, нездатною до пристосування; ймовірно, що в такій системі інформаційні канали, які ведуть до керуючого елементу, виявляться перевантаженими, а сам керуючий елемент буде нездатним опрацювати таку велику кількість інформації. Однак, чим децентралізованішими будуть рішення в системі, тим складніше їх узгодити з точки зору досягнення глобальної мети. Досягнення спільної мети в сильно децентралізованій системі може забезпечуватися лише стійким механізмом регулювання, який не дозволяє сильно відхилитися від поведінки, що веде до досягнення спільної мети. В системах, що не мають таких механізмів регулювання, наявність деякого рівня централізації є необхідністю. Загальне правило таке: *ступінь централізації повинен бути* ***мінімальним****, але таким, що забезпечить досягнення кінцевої мети*.

Системний аналіз у широкому розумінні – це метод пізнання, який використовує розкладання предмету дослідження на складові частини (елементи). Після цього дослідник намагається зрозуміти як взаємодіють ці елементи між собою. В цьому полягає синтез системи – об’єднання елементів у систему за допомогою зв’язків.

Системний аналіз використовує широкий математичний апарат та комп’ютерне програмне забезпечення для ефективного аналізу інформації про досліджуваний предмет в умовах невизначеності та неповноти даних.

Метою теорії систем та системного аналізу є відшукання принципів, загальних для різних складних об’єктів, на основі встановлення емпіричними дослідженнями їх ізоморфізму, функцій та динаміки.

Рис. 2.1. Етапи проведення системних досліджень [30, С. 38]

У загальному вигляді *системне дослідження проблеми* складається з таких етапів (рис. 2.1):

- формулювання проблеми;
- виявлення цілей;
- формулювання критеріїв;
- визначення ресурсів для досягнення цілей;
- генерація альтернатив та сценаріїв.

Більш детально, можна визначити наступні **основні етапи** системного аналізу.

1. Аналіз проблеми

Чи існує проблема? Точне формулювання проблеми. Аналіз логічної структури проблеми. Розвиток проблеми (у минулому і в майбутньому). Зовнішні зв'язки проблеми (з іншими проблемами). Принципова можливість розв'язання проблеми.

2. Визначення системи

Формулювання завдань, виходячи з проблеми. Визначення позиції спостерігача. Визначення об'єкта дослідження. Виділення елементів (визначення меж поділу системи). Визначення зовнішнього середовища.

3. Аналіз структури системи

Визначення рівнів ієрархії. Виділення підсистем. Визначення функціональних та структурних зв'язків.

4. Формулювання загальної мети і критерію системи

Визначення цілей системи – вимог надсистеми. Визначення обмежень середовища. Формулювання загальної мети. Визначення критеріїв. Декомпозиція критеріїв за підсистемами. Композиція загального критерію з критеріями підсистем.

5. Декомпозиція мети, визначення потреби в ресурсах

Формулювання макроцілей і тактичних цілей. Формулювання цілей підсистем. Виявлення потреб у ресурсах.

6. Виявлення ресурсів, композиція цілей

Оцінювання існуючої технології і потужностей. Оцінювання стану ресурсів. Оцінювання можливостей взаємодії з іншими системами. Оцінювання соціальних чинників. Композиція цілей.

7. Прогноз і аналіз майбутніх умов

Аналіз стійких тенденцій розвитку системи. Прогноз розвитку і зміни середовища. Передбачення виникнення нових чинників, що можуть впливати на розвиток системи. Аналіз майбутніх можливостей та ресурсів.

8. Оцінювання цілей і засобів

Оцінювання взаємозалежності і відносної важливості цілей. Оцінювання дефіцитності і вартості ресурсів. Оцінювання впливу зовнішніх чинників. Обчислення комплексних розрахункових оцінок за загальним критерієм.

9. Вибір альтернатив (варіантів)
Аналіз цілей на сумісність і повноту. Відсікання надлишкових цілей. Розроблення варіантів досягнення окремих цілей. Оцінювання і порівняння альтернатив. Синтез комплексу взаємозалежних варіантів.

10. Реалізація варіантів
Проектування організаційної структури та інформаційних механізмів. Виявлення недоліків організації управління та виробництва. Розробка математичної моделі, що імітує функціонування системи. Виявлення та аналіз заходів щодо удосконалення організації.

Системний аналіз об'єктів та процесів конкретизується в напрямку ***методологія–метод–нотація–засіб***. Деталізація методології здійснюється за допомогою методів, а відображення результатів застосування методів здійснюється за допомогою нотацій.

Методологія визначає основні керуючі положення для проведення системних досліджень і включає визначення понять, предметної області, принципи системного підходу, а також постановку та загальну характеристику основних проблем організації системних досліджень. При цьому, методологія для дослідження та розв'язання системних проблем сама має бути системною. Методологія використовується для оцінювання та обрання проекту системи, що розробляється, визначає кроки проектування, їх послідовність, правила вибору, розподілу та призначення методів.

Метод – це систематична процедура або техніка генерації описань компонентів системи (*наприклад*, проектування потоків та структур даних інформаційної системи).

До *основних методів системного аналізу* належать: метод аналізу ієрархій, метод дерева цілей, методи функціонального аналізу (функціонально-вартісний аналіз ФВА/ABC, аналіз ризику від дефектів FMEA, функціонально-фізичний аналіз ФФА, розгортання функцій якості QFD тощо), експертні методи (метод PATTERN, метод Дельфі тощо), методи комбінаторно-морфологічного аналізу і синтезу, евристичні методи генерування альтернатив (метод мозкового штурму, синектика, сценарний аналіз, ділові ігри тощо), системні методи на основі когнітивних карт, таблиць рішень та ін.

Нотації призначені для опису структури системи, елементів даних, етапів опрацювання. В набір нотацій включаються графи, діаграми, таблиці, блок-схеми, формальні та природні мови.

Засоби – це інструментарій для підтримання та посилення методів, вони підтримують роботу користувачів у процесі створення та редагування графічного проекту в інтерактивному режимі, сприяють організації проекту в вигляді ієрархії рівнів абстракції, реалізують перевірки компонентів на відповідність тощо.

2.2. Основні постулати загальної теорії систем

Загальна теорія систем (ЗТС) сформувалася в самостійний науковий напрямок у 30-х – 50-х роках XX ст. Засновниками ЗТС є біолог Л. Берталанфі, фахівець з математичних проблем в галузі біології Л. Хартлі, лінгвіст М. Месарович, бельгійський фізик І. Пригожин. Великий внесок у розробку ЗТС внесли і українські вчені, зокрема, батько української кібернетики В. М. Глушков.

Віктор Михайлович Глушков

Теоретичну базу загальної теорії систем складають кібернетика, теорія інформації, теорія ігор, теорія прийняття рішень, факторний аналіз, теорія графів, теорія мереж, теорія масового обслуговування та ін. До прикладної частини ЗТС належать системотехніка, дослідження операцій, ергономіка.

Загальна теорія систем вивчає всеможливі аспекти дослідження систем, зокрема, і прийняття рішень в них. За сучасними уявленнями до складу ЗТС входять такі науки як системний аналіз, кібернетика, інформатика, дослідження операцій та системотехніка.

Системний аналіз – це сукупність методологічних засобів, які використовуються для підвищення ступеня обґрунтованості рішень у складних слабко-структурованих і слабко-формалізованих проблемах. При цьому, системний аналіз передбачає розгляд об'єктів як систем, а основні методологічні засади системного аналізу базуються на принципах системного підходу. Розвиток методів системного аналізу, в основному, пов'язаний із розвитком складових частин загальної теорії систем.

Кібернетика – вивчає системи зі зворотним зв'язком і аспект керування інформацією в цих системах, розглядаючи при цьому строго формалізовані задачі.

Інформатика – займається дослідженням процесів збереження, накопичення, перетворення, передачі даних та інформації із застосуванням комп'ютерної техніки.

Дослідження операцій – вивчає методи прийняття рішень, розглядаючи при цьому переважно формалізовані задачі.

Системотехніка – наука, що вивчає застосування методів системного аналізу для дослідження технічних схем.

Основними напрямками системних досліджень за Берталанфі є наука про системи, системна технологія, системна філософія.

Наука про системи досліджує застосування системних концепцій у фізичних, суспільних науках та науках про поведінку емпіричним чином. Увага зосереджується на науковому вивченні цілого та цілісності, реалізуються оцінки рівнів складності та способів взаємодії між компонентами системи, для визначення подібності та ізоморфізмів використовуються математичні моделі.

Системна технологія розглядає проблеми у промисловості та суспільстві, які можна досліджувати шляхом застосування теорії систем. У системному аналізі, науці про управління, дослідженні операцій, інформатиці та промисловій інженерії концепції ЗТС конкретизуються для пошуку практичних розв'язань проблем.

Системна філософія намагається концептуалізувати взаємозв'язки та взаємозалежності між теоріями, сформульованими в різних сферах наукових досліджень, є спробою об'єднати різні розділи науки в межах філософських концепцій загальних систем.

Особливо слід відзначити розвиток системної технології стосовно технічних систем, що дозволило започаткувати новий науковий напрямок – системотехніку, що визначає методологію проектування складних технічних систем. ***Системотехніка*** виникла у США на початку 50-х років XX ст. і описує своєрідні «правила поведінки» інженера, що конструює складні системи.

Розвиток основних тенденцій в ЗТС спричинив виникнення ***системології***, котра розглядається як «комплекс понять і концепцій, що стосуються і системного підходу, і системного аналізу, і загальної теорії систем», тобто є поєднанням науки про системи та окремих аспектів системної філософії [9, С. 8–9].

Загальна теорія систем опирається на три ***постулати***.

1. Функціонування систем будь-якої природи може бути описане на основі розгляду структурно-функціональних зв'язків між окремими елементами систем.
2. Організація системи може бути визначена на основі спостережень за станом системи та її взаємодії з навколишнім середовищем.
3. Спосіб і рівень організації системи визначає ефективність її функціонування.

2.3. Види зв'язків між елементами системи

Як уже зазначалося, відносини між елементами системи реалізуються через зв'язки між ними. Зв'язки можуть бути енергетичними, речовинними, інформаційними, внутрішніми і зовнішніми, прямими і зворотними. Кожна система має вхід і вихід. На вхід поступають керуючі сигнали. З виходу надходять інформаційні сигнали (реакція системи). Через входи в систему надходить речовина, енергія чи інформація; результати процесів їхнього перетворення надходять у зовнішнє середовище через виходи.

Прямі зв'язки поділяються на: прямий *послідовний* зв'язок (a); *паралельний розподільчий* зв'язок (b); *паралельний з'єднуючий* зв'язок (c) (рис. 2.2).

Зворотний зв'язок – це зв'язок між виходом і входом системи. Він може здійснюватись безпосередньо або через інші елементи системи. Якщо зворотний зв'язок зменшує дію вхідного впливу на вихідну величину, його називають негативним; якщо зворотний зв'язок збільшує цей вплив – позитивним.

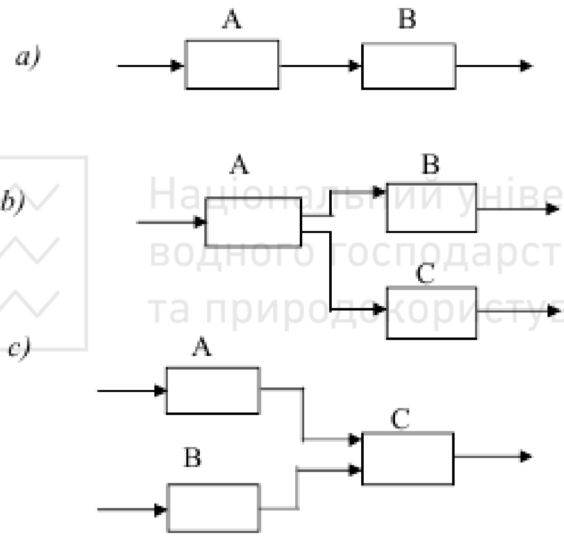


Рис. 2.2. Види прямих зв'язків у системах

Негативний зворотний зв'язок сприяє відновленню рівноваги в системі. Найпростішим прикладом такого зв'язку у технічних системах є регулювання рівня води у ванні через зливний отвір. Негативний зворотний зв'язок є важливим також для стабілізації економічних систем. Прикладом такого зв'язку є залежність між пропозицією товару та ціною.

Негативні зворотні зв'язки частково компенсують відхилення (диспропорцію), проте, не завжди повертають систему до початкових значень. Можуть встановитися нові значення цих параметрів, які є оптимальними для функціонування системи в нових умовах. В цьому і полягає принцип ***динамічної рівноваги***.

Позитивний зворотний зв'язок підсилює відхилення від рівноважного стану.

Приклад: банківський відсоток на депозити. При нарахуванні відсотків на відсотки (складні відсотки) сума грошей зростає все швидше і швидше.

При дії *позитивних зворотних зв'язків* деякі зміни в системі не пригнічуються, а нагромаджуються та підсилюються. Це може спричинити перехід системи до нерівноважного стану (система або руйнується, або перебудовується зі зміною структури та функцій). Позитивні зворотні зв'язки можуть відігравати як позитивну, так і негативну роль. Позитивний *приклад*: зі зростанням інвестицій зростає ефективність виробництва, а це сприяє залученню нових інвестицій, які стимулюють подальше зростання виробництва.

Прикладом небажаної дії позитивних зворотних зв'язків є механізм розгортання гіперінфляційної спіралі, коли в нерівноважному стані економіки зростання цін через механізм причинно-наслідкових зв'язків (панічні настрої, закупівлі про запас) спричиняє подальший виток зростання цін.

Наведені приклади показують, що економічні системи часто здатні самоорганізовуватися та налагоджувати взаємовідносини із середовищем, протидіяти негативним флуктуаціям та підтримувати гомеостатичний стан (динамічну рівновагу).

Аналіз елементів і зв'язків між ними дозволяє виявити структуру і функції системи. *Структура системи* – це внутрішня організація системи, яка відображає спосіб взаємодії утворюючих її компонентів. Для графічного зображення систем використовують структурні та функціональні схеми. Однією з найбільш поширених схем є ***блочна схема***, яка будується з блоків-елементів (прямокутників, паралелограмів, кругів), з'єднаних лініями чи стрілочками, які показують зв'язки чи потоки між ними (рис. 2.7). *Прикладом* блочної схеми є блок-схема алгоритму деякої комп'ютерної програми.

Іншою поширеною схемою системи є ***модель «чорний ящик»*** («чорна скринька»). Ця схема (рис. 2.3) використовується, коли дослідника не цікавить (чи йому не відома) внутрішня структура системи, а вивчається лише механізм її дії. Метод «чорного ящика» часто застовується для аналізу систем, структура яких невідома чи не береться до уваги, а відомі лише значення вхідних та вихідних параметрів. В такій моделі акцент робиться на призначенні та поведінці системи, а про її будову є тільки опосередкована інформація, що відображається у зв'язках із зовнішнім середовищем. Зв'язки з середовищем, що йдуть у систему (входи), дають можливість впливати на неї, використовувати її як засіб, а зв'язки, що йдуть із системи (виходи), є результатами її функціонування, які або впливають на зміни у середовищі, або споживаються зовні системи. Змінюючи вхідні сигнали та спостерігаючи за змінами вихідних сигналів можна вивчати механізм дії чорного ящика.

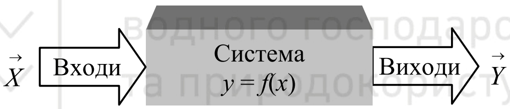


Рис. 2.3. Схема системи типу «чорний ящик» [30, С. 47]

Прикладами систем типу «чорний ящик» можна вважати автомобіль, комп'ютер чи смартфон для нефахівців, які у повсякденному житті користуються ними, не завжди знаючи досконало їх внутрішню будову та механізм дії.

Метод описування систем за допомогою моделі «чорний ящик» полягає у знаходженні взаємозв'язків між входами та виходами системи. Спостерігаючи достатньо довго за входами та виходами такої системи, тобто маючи вектори спостережень $\vec{X} = (x_1, ..., x_n)$ та $\vec{Y} = (y_1, ..., y_n)$, можна досягти такого рівня знань про її властивості, який дасть можливість передбачати зміни у вихідних компонентах при зміні вхідних, тобто можна знайти відображення $f : X \to Y$.

Для досягнення цієї мети будують спеціальні математичні моделі із застосуванням регресійного аналізу, математичної статистики, методів планування експериментів тощо. Необхідно зауважити, що дослідження системи методом «чорного ящика» принципово не може дати *однозначної інформації про її структуру*, бо однакову поведінку можуть мати різні системи.

Для детальнішого опису систем використовують ***моделі складу*** та ***моделі структури***. Модель складу системи відображає, з яких елементів та підсистем складається система (рис. 2.4), а модель структури застосовується для відображення відношень між елементами та зв'язків між ними.

Головна складність при побудові моделі складу полягає у тому, що поділ цілої системи на частини є відносним і залежить від мети дослідження. Крім того, відносним є поняття елемента– те, що з одного погляду є елементом, з іншого – може бути підсистемою. *Наприклад*, університет з точки зору його ректора, головного бухгалтера чи начальника служби охорони буде складатися із різних підсистем.

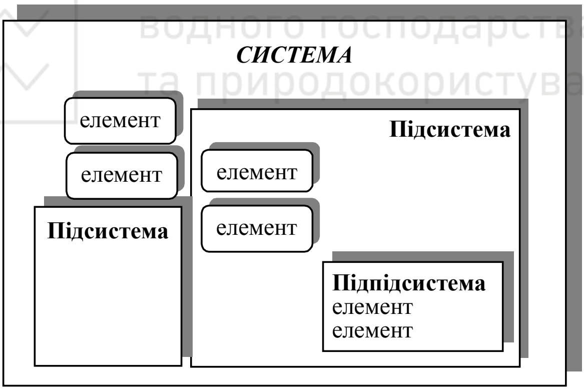


Рис. 2.4. Графічне зображення моделі складу системи [29, С. 48]

Модель структури описує суттєві зв'язки між елементами (компонентами моделі складу). *Прикладом структурної схеми* може бути ієрархічна структура управління корпорації (рис. 2.5).

Рис. 2.5. Структурна схема управління корпорації [30, С. 49]

2.4. Способи описання стану системи

Зміна стану реального об’єкта неминуче пов’язана з переносом речовини, енергії чи інформації, і не може відбуватись миттєво. Процес переходу з одного стану в інший називають перехідним процесом, а сам перехід – трансформацією стану.

Трансформація станів системи може бути подана у вигляді ***словесного описання***, ***логічної схеми***, ***графічного описання*** (схеми, графи) чи ***аналітичного описання*** (математичні вирази, системи рівнянь тощо).

Прикладом *словесного описання* може бути технологічна карта виконання робіт. *Наприклад,* карта робіт при облаштуванні шару дорожнього покриття з ґрунту, укріпленого бітумом, включає такі операції: транспортування ґрунту, розрівнювання ґрунту, транспортування води, зволоження ґрунту, транспортування бітуму, змішування ґрунту з бітумом, ущільнення покриття.

Приклад логічної схеми деякого підприємства подано на рис. 2.6.

Графічну схему можна розглядати як *графову модель* системи. Від графової моделі легко можна перейти до *аналітичного описання* системи.

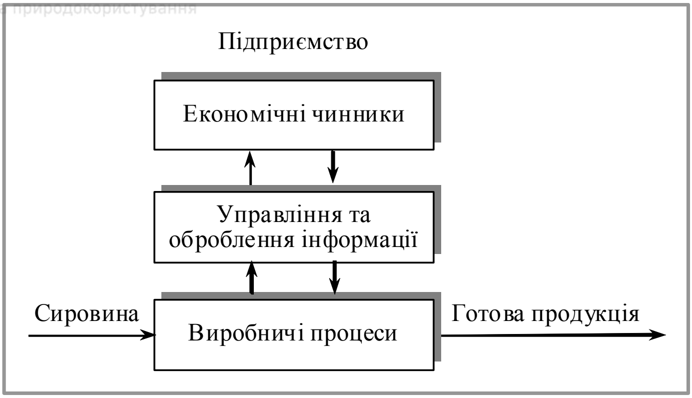


Рис. 2.6. Логічна схема підприємства [29, С. 46]

Приклад. Графова модель погодних станів. Погода розглядається як випадкова система, яка може перебувати в одному із трьох станів: S_1 – ясно, S_2 – похмуро, S_3 – йде дощ. Вважатимемо, що ймовірність стану погоди на наступний день залежить лише від сьогоднішньої погоди. На підставі статистичних даних встановлено, що матриця імовірностей переходів від одного погодного стану до іншого має вигляд

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}.$$

Подамо зміну погодних станів у вигляді графа, де стани зображено точками (кружками), а можливі переходи – стрілками (рис. 2.7). Числа біля стрілок позначають ймовірність відповідного переходу.

Графовій моделі можна поставити у відповідність *аналітичну математичну модель*. Вона матиме вигляд системи трьох лінійних рівнянь, де у лівій частині кожного з рівнянь стоїть сума добутків перехідних ймовірностей на ймовірності попередніх станів, у правій частині – ймовірності наступних станів

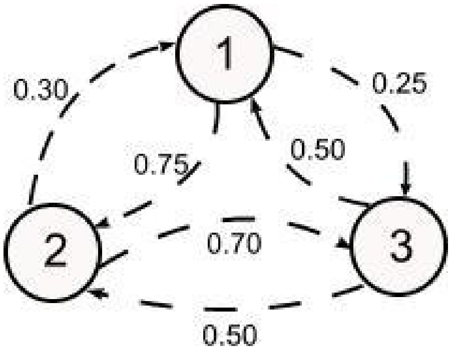


Рис. 2.7. Графова модель погодних станів

Розв'язавши систему (2.1), знайдемо середні ймовірності стаціонарних станів системи, тобто таких станів, які зустрічаються найчастіше (стани динамічної рівноваги).

$$\begin{cases} 0{,}30\,p_2 + 0{,}50\,p_3 = p_1; \\ 0{,}75\,p_1 + 0{,}50\,p_3 = p_2; \\ 0{,}25\,p_1 + 0{,}70\,p_2 = p_3. \end{cases} \qquad (2.1)$$

Складну систему, як правило, неможливо «охопити» повністю та детально описати, що на практиці не завжди й потрібно. Основна проблема при описанні систем полягає у тому, що доводиться знаходити компроміс між простотою описання та необхідністю врахування численних чинників та характеристик складної системи. Як правило, цю проблему вирішують через описання системи не однією моделлю, а кількома чи сімейством моделей, кожна з яких описує поведінку системи при різних рівнях абстрагування. Для кожного рівня існує низка характерних особливостей і змінних, законів та принципів, за допомогою яких описується поведінка системи. Для того, щоб таке описання (його ще називають ієрархічним) було ефективним, необхідно розробляти якомога більшу кількість незалежних моделей для різних рівнів системи.

Автономність цих моделей, проте, не виключає, що кожна модель має певні зв'язки з іншими.

Процес поділу системи на рівні, що характеризують технологічні, інформаційні, економічні та інші аспекти її функціонування, називають ***стратифікацією*** системи, а самі рівні – ***стратами***. На кожній страті в ієрархії структур є свій власний набір змінних, які дають змогу значною мірою обмежитись лише дослідженням одного аспекту системи, однієї страти. Незалежність страт дає можливість глибше та детальніше досліджувати системи, хоча припущення про їх незалежність може призвести до неповного розуміння поведінки системи в цілому.

Приклади. Для описування функціонування інформаційної системи використовують, як правило, не менше двох моделей. Перша описує систему мовою фізичних законів, що управляють функціонуванням та взаємодією її складових. Друга подає абстрактні логічні поняття, такі як файли, інформаційні потоки тощо. На рівні фізичних процесів розглядається правильне функціонування різноманітних електронних пристроїв (компонентів), а на логічному рівні – проблеми оброблення інформації, обчислення, програмування тощо.

Виробничий комплекс (рис. 2.6) моделюють, як правило, на трьох рівнях: виробничому (фізичні процеси випуску продукції), управління та оброблення інформації, економічному (з погляду продуктивності та прибутковості). Для кожного з цих трьох аспектів системи існує своя мова описання, свої моделі, хоча система залишається тією самою.

Теми для роздумів і контрольні запитання

1. Що відображають принципи системного підходу?
2. Яким чином можна застосувати принципи системного підходу на практиці?
3. Яким має бути рівень централізації для ефективного функціонування системи?
4. Що таке системний аналіз?
5. Етапи проведення системних досліджень.
6. Роз'ясніть суть понять методологія, метод, нотація, засіб.

7. Поняття про загальну теорію систем. Які основні тенденції в ЗТС були визначені Людвігом фон Берталанфі?
8. Постулати теорії систем.
9. Які проблеми розглядає системна технологія?
10. Що є предметом системної філософії?
11. Що описує системотехніка?
12. Види зв'язків між елементами системи.
13. Наведіть приклади негативних та позитивних зворотних зв'язків в системі.
14. Яку роль відіграють інформаційні потоки в складних системах?
15. Модель складу і модель структури. Наведіть приклади.
16. Схема системи типу «чорний ящик» і її роль у кібернетиці.
17. Переваги та недоліки моделі системи типу «чорний ящик».
18. Способи опису стану системи.
19. Наведіть приклади задання стану системи за допомогою словесного, графічного, аналітичного описання**,** логічної схеми**.**
20. Поняття про стратифікацію системи. Наведіть приклади.

3. МЕТОДИ ПРЕДСТАВЛЕННЯ ДАНИХ

3.1. Класифікація вимірювальних шкал

Для визначення стану системи необхідно оцінити (виміряти) значення її основних параметрів. В основі аналізу систем та прийняття рішень лежать вимірювання, що виконуються з метою конструювання та дослідження систем, дозволяють зробити висновки про те, чи краща одна з альтернатив за іншу, оцінити «силу» переваги.

Вимірювання – це присвоєння об'єктам чи окремим їх властивостям числових або нечислових характеристик за певними правилами. Позначимо через $x_i, \left(i = \overline{1,m}\right)$ спостережені властивості системи, а через $y_j,\ j = \left(\overline{1,m}\right)$ – значення (позначення) цих властивостей. Чим тіснішою є відповідність між станами та їх значеннями, тим більше інформації можна отримати в результаті обробки таких даних.

В теорії вимірювань можна виділити дві основні проблеми:

- визначення типу шкали вимірювань для конкретних даних;
- пошук алгоритмів аналізу даних, результат яких не змінюється за будь-якого допустимого перетворення шкали (тобто дані є інваріантними щодо цього перетворення).

Множина значень, що використовується для реєстрації станів спостережуваного об'єкта, називається **вимірювальною шкалою**. Вимірювальні шкали, в залежності від допустимих на них операцій, розрізняються за силою. Найслабшими є номінальні шкали, а найсильнішими – абсолютні. Виділяють три основні атрибути вимірювальних шкал, наявність або відсутність яких визначає приналежність шкали до тієї чи іншої категорії:

1) ***впорядкованість*** означає, що деяке значення шкали є більшим, меншим або рівним за інше значення;

2) ***інтервальність*** означає, що інтервал між будь-якою парою значень шкали є більшим, меншим або рівним за інтервал між іншою парою значень;

3) ***нульова точка*** (точка відліку) означає, що набір значень відповідних до вимірюваних властивостей має точку відліку, яка відповідає повній відсутності вимірюваної властивості.

У відповідності з цим виділяють наступні *види шкал*:

- ***неметричні*** або якісні шкали, в яких відсутні одиниці вимірювань (номінальна і порядкова шкали); номінальна шкала не має властивостей впорядкованості інтервальності та точки відліку, порядкова шкала має лише впорядкованість;
- ***кількісні*** або ***метричні*** (присутня шкала інтервалів, шкала відношень, абсолютна шкала).

Шкала найменувань

Шкала найменувань (*номінальна шкала*) містить скінченний набір значень (позначень) для різних станів об’єкта. При цьому відсутні впорядкованість, інтервальність, нульова точка. Вимірювання полягає в тому, щоб визначити належність результату до того чи іншого стану і записати це за допомогою символу або слова, що позначає такий стан. Це найпростіша шкала, яка використовується лише з метою відрізнити один об’єкт від іншого.

Якщо класифікуються дискретні за своєю природою об’єкти чи явища, які не мають кількісних характеристик, а лише якісні, то найприродніше використовувати саме шкалу найменувань (яку називають часто також *класифікаційною*). Ця шкала класифікує об’єкти чи окремі їхні ознаки для їх розпізнавання та виявлення подібності або відмінності. Числові значення використовуються як назви (імена). Єдина функція цієї шкали – розрізнення об’єктів.

Приклади:

- географічні назви, власні імена людей тощо;
- герби та прапори держав, емблеми родів військ, всілякі значки тощо;

- номери автомобілів, офіційних документів, номери на майках спортсменів;
- поштові адреси, печатки тощо.

Необхідно розуміти, що позначення класів – це лише символи, з якими не можна діяти як з числами. *Приклад*. Якщо в одного спортсмена на спині номер 1, а іншого – 2, то не можна сказати, що «другий в два рази сильніший за першого».

З даними, представленими номінальною шкалою, можна виконувати тільки операцію перевірки їх співпадіння чи розбіжності.

Порядкові шкали

Наступна за силою після номінальної шкали – порядкова (*рангова*) шкала. Вона застосовується в тих випадках, коли вимірювана ознака дозволяє порівнювати різні об'єкти. Ця шкала дозволяє ранжувати об'єкти чи сукупності їхніх ознак за пріоритетом. Числа в цих шкалах відображають порядок розміщення елементів – ранги об'єктів (або їхніх ознак) за пріоритетом. При цьому не можна визначити міру домінування, тобто виміряти, наскільки один об'єкт кращий чи важливіший за інший. Тобто присутня упорядкованість, але відсутні атрибути інтервальності та нульової точки. Єдиними типами відношень між значеннями порядкової шкали можуть бути:

а) рівність або нерівність значень порядкових змінних величин;
б) відношення «більше» або «менше» між різними значеннями порядкових ознак.

Вимірювання в шкалі порядку може застосовуватися, *наприклад*, у таких ситуаціях:

- коли необхідно впорядкувати об'єкти в часі або просторі;
- коли потрібно впорядкувати об'єкти відповідно до якоїсь характеристики.

Характерною особливістю порядкових шкал є те, що в них встановлюється тільки відношення порядку, але нічого не говориться про те, <u>наскільки</u> один клас об'єктів відрізняється (кращий чи гірший) від іншого. Тому у випадку, коли порядок встановлюється певними числами, немає сенсу говорити про середнє число для певної групи чи про ступінь переваги однієї групи над іншою.

Наприклад, порядковою шкалою користуються при оцінці виступів гімнастів, оцінці фігурного катання на ковзанах. Зрозуміло, що спортсмен, який одержав 10 балів, виступив краще за того, хто одержав 5 балів, проте не можна говорити, що він виступив в 2 рази краще від свого суперника.

Типові порядкові шкали. Позначивши класи символами й установивши між цими символами відносини порядку, ми отримаємо шкалу простого порядку: A → B → C → D → E → F. *Приклади*: нумерація в журналі, телефонний довідник.

Іноді виявляється, що деякі пари не можна впорядкувати і вони вважаються рівними A ≥ B і B ≤ A, тобто A = B. Відповідна шкала називається ***шкалою слабкого порядку***.

Приклад такої шкали – упорядкування за ступенем близькості з конкретною людиною («мати» = «батько» > «син» = «дочка», «дядько» = «тітка» < «брат» = «сестра» тощо).

Інша ситуація виникає, коли є пари класів, незрівнянні між собою, тобто ні A ≥ B, ні B ≤ A. У такому випадку говорять про ***шкалу часткового порядку***. Шкали часткового порядку часто виникають в соціологічних дослідженнях.

Наприклад, при вивченні купівельного попиту суб'єкт часто не в змозі оцінити, який саме з двох товарів йому більше подобається (смартфон чи планшет); підліток не може вибрати улюблене заняття (футбол чи музика).

Модифіковані порядкові шкали. Часто використовують модифіковані і частково підсилені порядкові шкали.

1. *Шкала Мооса*. У 1811 р. німецький мінералог Ф. Моос запропонував встановити ***шкалу твердості мінералів***, постулюючи тільки десять її градацій. За еталони прийнято такі мінерали за зростанням твердості: 1 – тальк, 2 – гіпс, 3 – кальцій, 4 – флюорит, 5 – апатит, 6 – ортоклаз, 7 – кварц, 8 – топаз, 9 – корунд, 10 – алмаз. З двох мінералів твердішим є той, який залишає на іншому подряпини або вм'ятини при досить сильному зіткненні. Однак номери градацій алмазу і апатиту не дають підстав стверджувати, що алмаз є в два рази твердішим за апатит.

2. ***Шкала сили вітру за Ботфортом.*** У 1806 р. Ф. Ботфорт запропонував умовну 12–бальну шкалу для оцінки сили вітру за його дією на наземні предмети та за хвилюванням моря: 0 – штиль (затишшя), 4 – помірний вітер, 6 – сильний вітер, 10 – буря (шторм), 12 балів – ураган.

3. ***Шкала магнітуд землетрусів за Ріхтером.*** Американський сейсмолог Ч. Ріхтер у 1935 р. запропонував 12–бальну шкалу землетрусів за магнітудами, що базується на оцінці енергії сейсмічних хвиль, які виникають під час землетрусів.

4. ***Бальні шкали оцінки знань учнів, студентів.***

Для оцінювання результатів навчання учнів (студентів) використовуються такі шкали: 5–бальна (національна шкала у закладах вищої освіти), 12–бальна (шкала у закладах середньої освіти), 100–бальна (шкала ЄКТС для закладів вищої освіти), 200–бальна (шкала для оцінювання результатів ЗНО) тощо.

Числові шкали

Інтервальні шкали. Наступна за силою – інтервальна шкала, яка на відміну від попередніх, є кількісною шкалою. Шкали інтервалів використовуються тоді, коли однакові різниці числових значень, виміряні в них, відповідають однаковим різницям вимірюваної ознаки. Вони дають змогу виміряти «відстань» між об'єктами, визначити, на скільки одиниць виміру одна властивість краща чи гірша за іншу. В інтервальній шкалі присутні упорядкованість та інтервальність, але немає нульової точки. Різні шкали для вимірювання однієї й тієї ж характеристики можуть мати різні нульові точки відліку. Для інтервальної шкали початок відліку або відсутній, або вибирається довільно чи умовно.

Прикладами величин, що вимірюються інтервальними шкалами є температура, час, висота місцевості – тобто, ті величини, які за фізичною природою або не мають абсолютного нуля, або допускають свободу вибору у встановленні початку відліку.

Розглянемо ці приклади детальніше. При вимірюванні висоти місцевості чи підйому в гори говорять про висоту над рівнем моря. Проте, якого моря? Адже рівень морів та океанів різний і змінюється з часом. Існують різні системи відліку. В Україні висоти точок земної поверхні відраховують за Балтійською системою висот (від середнього рівня Балтійського моря в районі Кронштадту), в Європі – від рівня Середземного моря, в Росії і у Східній Азії використовують тихоокеанську систему висот.

Для вимірювання температури також використовують різні шкали, зокрема, шкалу Цельсія і шкалу Фаренгейта. В температурній шкалі шведського вченого Андерса Цельсія, запропонованій у 1742 році, за нуль (0° C) прийнято температуру плавлення льоду (замерзання води), а за 100° C – температуру кипіння води (конденсації пари). Фізик Данієль Габріель Фаренгейт у шкалі, запропонованій у 1724 році, за нульову точку (0° F або −17.78° C) прийняв температуру соляного розчину (лід, вода і хлорид амонію в співвідношенні 1:1:1), а за другу точку 32° F – точку плавлення льоду (0° C). Попри те, що між цими шкалами є взаємно-однозначна відповідність (див. формулу (1.3)), не можна сказати, що температура води збільшилася в два рази при її нагріванні від 10° C до 20° C за шкалою Цельсія, оскільки за шкалою Фаренгейта температура в тому ж досліді зміниться з 50° F до 68° F.

Для вимірювання часу використовують, так званий, всесвітній час, відлік у якому ведеться від часу за Гри́нвічем (англ. Greenwich Mean Time) – часу середньої півночі на меридіані, що проходить через Грінвіцьку королівську обсерваторію (поблизу Лондона). Для визначення місцевого часу необхідно враховувати часовий пояс, тобто, коли в Гринвіцькому часовому поясі (наприклад, на Острові Святої Єлени) час 21:28:10, то у Києві (у часовому поясі на дві години на схід від Гринвіча) – час 23:28:10.

Людство в різні епохи користувалось різними календарями. Найбільш відомими з них є: *Григоріанський* – літочислення ведеться від дня народження Ісуса Христа (нова ера); *Юліанський* – літочислення ведеться від гіпотетичної дати створення світу 5506 року до народження Ісуса Христа (до нової ери); *Іудейський* – літочислення від створення Адама – 3696 рік до народження Ісуса Христа (до нової ери); *Магометанський* – літочислення від втечі Магомета з Мекки в Медину – 622 рік після народження Ісуса Христа (нової ери).

В інтервальній шкалі можна виконувати арифметичні операції над інтервалами, але не можна виконувати операції над самими відліками на шкалі.

Циклічні (періодичні) шкали. Частковим випадком інтервальних шкал є циклічні шкали. У такій шкалі значення не змінюється при додаванні деякого числа (періоду). Для даних, виміряних у цій шкалі, можна застосовувати такі арифметичні дії як додавання і віднімання.

Приклади: шкала компаса (період – кут 360 град), циферблат годинника (періоди – 12 год, 60 хв).

Шкали відношень. Наступною за силою шкалою є шкала відношень або *метрична пропорційна* шкала. Для вимірювань такої шкали можна виконувати всі арифметичні дії. Тут присутні всі атрибути вимірювальних шкал: упорядкованість, інтервальність, нульова точка. Ці шкали найпоширеніші серед кількісних шкал у науці і практиці. У шкалі відношень вимірюють більшість фізичних величин, які мають природну нульову точку відліку: масу тіла, довжину, електричний заряд тощо. Допустимими для шкал відношень є перетворення подібності (ті, що змінюють лише масштаб).

Приклади: вага, довжина, гроші – величини, природа яких відповідає шкалі відношень. Порівнюючи значення за шкалою відношень бачимо, у скільки разів властивість одного об’єкта перевершує таку ж властивість іншого об’єкта.

Абсолютна (*метрична*) шкала має абсолютний нуль та абсолютну одиницю. У цій шкалі числові значення задано з точністю до тотожних перетворень. В абсолютній шкалі фіксованими є і початок відліку і масштаб.

Прикладами абсолютних шкал можуть бути: числова вісь, шкала температур за Кельвіном.

Середні величини

Середні значення в системному аналізі, теорії управління та теорії прийняття рішень використовують, зазвичай, для заміни сукупності чисел (кількісних значень) одним узагальненим (або згенерованим) показником, тобто для порівняння різних сукупностей шляхом порівняння їх середніх значень. Іншими словами, середнє значення сукупності є частковим випадком інформаційного фільтру при опрацюванні інформації.

На практиці використовують різні варіанти середньої величини: моду, медіану, середнє арифметичне, середнє геометричне, середнє гармонічне, середнє квадратичне тощо. Для значень, поданих у шкалі найменувань, середнього значення не існує. У порядковій шкалі як середнє можна використовувати моду або медіану, але середнє арифметичне чи середнє геометричне обчислювати недоцільно. У числових шкалах можна використовувати різні види середніх величин, в залежності від потреб конкретної задачі, зокрема, в шкалі інтервалів найдоцільніше застосовувати саме середнє арифметичне.

Мода (*Мо*) − це варіант ознаки, який найчастіше зустрічається у варіаційному ряду.

Наприклад, ціна, за якою найчастіше реалізується товар на ринку, є модою або модальною ціною; місячна заробітна плата, яка найчастіше зустрічається у даному колективі, є для нього модальною заробітною платою.

На практиці моду знаходять, як правило, за згрупованими даними. У дискретному ряду мода визначається без обчислення як значення ознаки з найбільшою частотою.

Медіана (*Ме*) − середній показник у групі впорядкованих показників. Половина показників мають значення більші, ніж медіана, інша половина показників є меншими від медіани.

Наприклад, медіана сукупності (2, 3, 3, 4, 5, 7, 9) становить 4. Мода цієї сукупності дорівнює 3. Середнє арифметичне сукупності дорівнює 4,71.

Загальні формули розрахунку статистичних середніх мають показник степеня (m). У залежності від того, яке значення він приймає, розрізняють такі види середніх:

- середнє гармонійне $m = -1,\quad \bar{x} = \dfrac{n}{\sum \frac{1}{x}}$;
- середнє геометричне $m \to 0,\quad \bar{x} = \sqrt[n]{\prod x}$;
- середнє арифметичне $m = 1,\quad \bar{x} = \dfrac{\sum x}{n}$;
- середнє квадратичне $m = 2,\quad \bar{x} = \sqrt{\dfrac{\sum x^2}{n}}$;
- середнє кубічне $m = 3,\quad \bar{x} = \sqrt[3]{\dfrac{\sum x^3}{n}}$,

де n позначає кількість елементів сукупності; $\sum x$ – суму всіх елементів x, $\prod x$ – добуток всіх елементів x.

Якщо розрахувати різні види середніх для одних і тих же вихідних даних, то значення їх виявляться неоднаковими. Тут діє *правило мажорантності*: із збільшенням показника степеня m збільшується і значення відповідної середньої величини

$$\bar{x}_{гарм} \le \bar{x}_{геом} \le \bar{x}_{арифм} \le \bar{x}_{квадр} \le \bar{x}_{куб}.$$

Розглянемо жартівливий *приклад*, як користуючись правилом мажорантності статистика може залежно від настрою і бажання її знавця «втопити» або «виручити» студента, який отримав на сесії оцінки 2 і 5. Який його середній бал? Якщо судити за середнім арифметичним, то середній бал дорівнює 3,5

$$\bar{x}_{арифм} = \frac{2+5}{2} = 3{,}5.$$

Але якщо декан бажає «втопити» нещасного і обчислить середнє гармонійне

$$\bar{x}_{гарм} = \frac{2}{1/2 + 1/5} = 2{,}86,$$

то студент залишається і в середньому двієчником, не дотягнувши до трійки.

Проте студентський комітет може заперечити декану і представити середню кубічну величину

$$\bar{x}_{куб} = \sqrt[3]{\frac{2^3 + 5^3}{2}} = \sqrt[3]{66{,}5} \approx 4{,}05.$$

Студент вже виглядає «хорошистом» і навіть претендує на стипендію. І тільки в тому випадку, коли «ледар» провалив обидва іспити, статистика допомогти не в змозі: на жаль, всі середні значення з двох двійок дорівнюють все тій же двійці!

Природна система аксіом (вимог до середніх величин) приводить до поняття асоціативних середніх, одним із яких є «середнє за Колмогоровим» (1930 р.), що обчислюється за формулою

$$\bar{x} = G\left(\frac{1}{n}\sum_{i=1}^{n} F(x_i)\right),$$

де F – строго монотонна функція (строго спадна або строго зростаюча), $G = F^{-1}$ – обернена функція до F.

«Середнє за Колмогоровим» − окремий випадок середнього, що узагальнює декілька видів:

- якщо $F(x) = x$, то отримаємо середнє арифметичне;
- якщо $F(x) = ln\, x$, то отримаємо середнє геометричне;

- якщо $F(x)=\frac{1}{x}$, то отримаємо середнє гармонічне;
- якщо $F(x)=x^2$, то отримаємо середнє квадратичне.

Моду і медіану не можна подати як «середнє за Колмогоровим».

Оскільки в теорії управління та прийняття рішень необхідно застосовувати лише інваріантні алгоритми опрацювання даних, то окрім середніх величин, використовуються й інші статистичні характеристики: показники розсіювання, зв'язку, відстані та ін. Так, коефіцієнт кореляції не змінюється за будь-якого допустимого перетворення в шкалі інтервалів, відношення дисперсії – у шкалі різниць, коефіцієнт варіації – у шкалі відношень тощо.

3.2. Проблеми, які виникають при обробці даних

При проведенні вимірювань та опрацюванні даних часто стикаються з поняттям невизначеності. ***Невизначеність*** – це властивість об'єкта (системи), що виражається в його незрозумілості, неясності, яка приводить до недостатньої можливості аналізу, розуміння, визначення його теперішнього чи майбутнього стану.

Найважливіші види невизначеності, що виникають у системному аналізі та теорії прийняття рішень, можна класифікувати так.

I. Невідомість
II. Недостовірність
   1. Неповнота
   2. Недостатність
   3. Недовизначеність
   4. Неадекватність
III. Неоднозначність
   1. Фізична невизначеність
      А. Випадковість
      Б. Неточність
   2. Лінгвістична невизначеність
      А. Невизначеність значень слів (полісемія)

— Омонімія
— Нечіткість

Б. Невизначеність змісту фраз

— Прагматична
— Семантична
— Синтаксична

3. Невизначеність мети (багатокритеріальність)
4. Багатоособовість

На першому рівні розташовані основні чинники, від яких залежить вид невизначеності: невідомість, недостовірність і неоднозначність. У ситуації *невідомості* фактично відсутня інформація про задачу; здебільшого це буває на початковій стадії дослідження. Якщо в процесі збирання інформації на певному етапі виявляється, що зібрано не всю інформацію чи одержати її з певних причин неможливо, то невизначеність трансформується в *недостовірність*. Вона може набирати вигляду *неповноти* чи *недостатності* (є не вся потрібна інформація), для деяких задач є неточні описи (*недовизначеність*), певні елементи задачі описано лише за аналогією з уже розв'язуваними (*неадекватність*).

Причини можливої неоднозначності опису – зовнішнє середовище (*фізична невизначеність*) і фахова мова, що використовується аналітиком (*лінгвістична невизначеність*).

Фізична невизначеність може бути пов'язана як із наявністю в зовнішньому середовищі кількох можливостей, кожна з яких реалізовується випадково (ситуація *випадковості* або стохастичної невизначеності), так і з неточністю вимірювань величини за допомогою фізичних приладів (ситуація *неточності*).

Лінгвістична невизначеність виникає внаслідок використання природної мови (в окремому випадку – фахової мови аналітика) для описання задачі. Цей вид невизначеності зумовлений необхідністю оперувати скінченною кількістю слів і обмеженим набором структурних фраз (речень, абзаців, текстів) для описання за скінченний час нескінченної множини різноманітних ситуацій, що виникають.

Лінгвістична невизначеність породжена, з одного боку, множинністю значень слів (понять і відношень) мови (полісемією), а з іншого – неоднозначністю змісту фраз. Виділяють два види полісемії: омонімію та нечіткість. Якщо об'єкти задачі, що відображаються одним і тим же словом, суттєво різняться, то така ситуація належить до омонімії, *наприклад*: коса – вид узбережжя, сільськогосподарський інструмент, зачіска. Коли ж ці об'єкти подібні, то це нечіткість, *наприклад*: невеликий запас пального на складі – 1.0 т, 1.1 т тощо; множина чисел, значно менших за тисячу.

Щодо джерел неоднозначності змісту фраз, вирізняють синтаксичну, семантичну та прагматичну неоднозначність. У першому випадку, уточнивши синтаксис, можна зрозуміти зміст фрази, *наприклад*, «залізні болти та гайки» – болти залізні, а гайки можуть бути з іншого металу; або і болти і гайки залізні; «стратити не можна помилувати» – «стратити не можна, помилувати» або ж «стратити, не можна помилувати». У випадку семантичної невизначеності окремі слова зрозумілі, але неясний зміст усієї фрази, наприклад: «блакитні зелені думки люто сплять». Прагматична невизначеність спричинена неоднозначністю використання інформації для досягнення певних цілей.

Невизначеність може виникати й через невизначеність мети (це веде до виникнення задач із багатьма критеріями), а також у багатоособових задачах, коли є кілька експертів чи осіб, що приймають рішення. У випадку активної протидії в одних ситуаціях чи активного сприяння в інших така невизначеність моделюється методами теорії ігор.

Урахування фізичної невизначеності може ускладнитися виникненням лінгвістичної невизначеності в описанні розподілу ймовірностей. Інакше кажучи, різні види невизначеності можуть накладатися одна на одну.

З погляду часу вирізняють *перспективну* невизначеність (виникають непередбачені чинники) та *ретроспективну* (брак інформації про поведінку об'єкта в минулому). У випадку ретроспективної невизначеності можливі три варіанти:

інформацію можна відновити, можна замінити перспективною, не можна ні відновити, ні замінити.

Розглянемо більш детально два види невизначеності, які формалізуються методами математичного моделювання: випадковість і нечіткість. ***Випадковість*** – це вид невизначеності, який підкоряється лише закону розподілу ймовірностей. Знаючи розподіл імовірностей $p(x)$, можна відповісти на більшість запитань про випадкову величину: якому інтервалу належать її можливі значення; навколо якого значення вони розсіюються (середнє вибіркове); наскільки сильно розкидані ці значення (дисперсія) тощо. Для цього достатньо знати не весь розподіл, а лише основні його параметри (середнє вибіркове, дисперсію).

Стохастична модель системи чи процесу відображає зв'язки між вимірними величинами (залежності ендогенних змінних від низки інших ендогенних змінних і/чи екзогенних змінних), причому хоча б деякі з цих зв'язків мають імовірнісний характер і/чи хоча б деякі з величин є випадковими. Самі зв'язки формалізуються як рівняння, нерівності чи умови оптимізації функцій ендогенних змінних. Стохастичні динамічні моделі враховують випадковість та невизначеність, що притаманні динамічним системам і дозволяють більш адекватно описати існуючі складні соціально-економічні чи інформаційні процеси. Появі стохастичних моделей сприяло нагромадження статистичних даних, що дало дослідникам можливість працювати не лише з абстрактними моделями, але й з реальними параметрами і результатами. З'явилась можливість не тільки корегувати теоретичні моделі, але й враховувати чинники, що мають випадковий характер. У рівняннях стохастичних моделей відсутня жорстка функціональна залежність між параметрами (змінними), або функціональний зв'язок враховує випадкові події.

Стохастичні динамічні моделі часто передбачають використання регресійних рівнянь для моделювання відповідних процесів. Генерація стохастичних регресійних рівнянь, як правило, здійснюється на основі часових рядів або табличних ретроспективних даних за допомогою класичного методу найменших квадратів і автокореляційної функції першого порядку. При цьому поряд із звичайними змінними використовуються лагові та фіктивні змінні [6, С. 77–84].

Під *випадковим (стохастичним) процесом* (випадковою функцією часу) розуміють функцію $x(t)$, котра може мати ту чи іншу конкретну реалізацію (траєкторію) з деякої фіксованої множини можливих траєкторій. Таким чином, в умовах невизначеності моделлю динаміки стану системи може виступати векторний випадковий процес $\tilde{x}(t)=(\tilde{x}_1(t),\ldots,\tilde{x}_n(t))$, кожна компонента $\tilde{x}_j(t)$ якого описує стохастичну динаміку j-ї $\left(j=\overline{1,n}\right)$ характеристики системи.

Іншим видом невизначеності є ***нечіткість*** даних. Найчастіше ситуація нечіткості зустрічається при користуванні мовними (лінгвістичними) конструкціями. Так, у виразі: «Висока молода людина» – названо клас, до якого належить людина, але невідомо, який у неї зріст та скільки їй років. Інші приклади нечіткості: «гарячий», «дешевий», «розумний».

Лінгвістична змінна – це змінна, значення якої розпливчасте за своєю природою. Для операцій з лінгвістичними змінними створено математичний апарат – теорію ***нечітких множин***.

Ознаки, за якими елементи включають до нечіткої множини, не дають змоги однозначно відокремити ті елементи, що входять до неї, від тих, що не належать їй. Для встановлення належності об'єкта до нечіткої множини використовують поняття **функції належності** $\mu_A(x)$. Для кожного значення змінної x можна задати число $\mu_A(x)$, $0\le\mu_A(x)\le 1$, яке виражає ступінь належності цього значення до нечіткої множини A. Якщо $\mu_A(x)=0$, то елемент x не належить множині A, якщо $\mu_A(x)=1$ – однозначно належить.

Нечітку множину A визначають як сукупність упорядкованих пар вигляду:

$$A=\{x,\mu_A(x)\},\ x\in X\ . \qquad (3.1)$$

Приклади: нечітку множину «студент добре вчиться» можна задати у вигляді $\{(2, 0.0); (3, 0.2); (4, 0.8); (5, 1.0)\}$, де перший елемент пари – оцінка студента, другий – значення функції належності до множини;

нечітку множину «теплий день» – можна задати у вигляді $\{(15, 0.4); (20, 0.7); (25, 1.0); (30, 0.4); (35, 0.1)\}$, де перший елемент – температура повітря, другий – значення функції належності до множини.

Окрім невизначеності **при обробці даних виникають і інші проблеми.**

Велика розмірність. У багатьох дослідженнях кількість об'єктів N та кількість їх ознак n великі, тому добуток $N \times n$ має декілька десяткових порядків. Врахування часу призводить до ще більшого зростання розмірності блоку даних $N \times n \times t$. Застосування ПК істотно розширює можливості обробки даних, але «прокляття розмірності» залишається серйозною проблемою.

Різнотипність даних. Різні ознаки вимірюють в різних шкалах та одиницях. Виникає необхідність їх порівняти. *Наприклад*, є наступні дані, які описують роботу підприємства: прибуток (1,5 млн. грн.), рентабельність (10%), об'єм використаної води (160 куб. м). Необхідно побудувати інтегральний показник, який характеризує ефективність роботи підприємства. Для цього потрібно звести дані до однієї шкали та розмірності, або ж побудувати алгоритм обробки різнотипних даних.

Пропущені значення. Часто зустрічаються незаповнені комірки таблиці даних (рівні часового ряду). Найпростішим способом відновлення пропущеного значення є обчислення середнього між попереднім та наступним значеннями у часовій шкалі.

Зашумленість. Зазвичай результати вимірювання відрізняються від фактичних значень на деяку випадкову величину – похибку. Якщо статистичні властивості похибок не залежать від вимірюваної величини, таку похибку називають

адитивним шумом. Для очищення даних від похибок (видалення шуму) здійснюють фільтрування сигналів.

Теми для роздумів і контрольні запитання

1. Поняття вимірювальної шкали, її атрибути, «сила» шкали.
2. Для чого використовується шкала найменувань (номінальна шкала)? Наведіть приклади.
3. Що означають числа в порядковій (ранговій) шкалі?
4. Які дії можна виконувати в інтервальній шкалі?
5. Наведіть приклади метричних шкал.
6. Які Ви знаєте середні величини?
7. Які середні величини можна використовувати у порядковій шкалі, у інтервальній шкалі, у метричній шкалі?
8. Поняття невизначеності при обробці даних. Види невизначеності.
9. Яким чином моделюються випадкові значення та процеси?
10. Поняття нечіткості. Лінгвістична змінна.
11. Що таке нечітка множина? Наведіть приклади.

4. ІНФОРМАЦІЙНІ АСПЕКТИ ВИВЧЕННЯ СИСТЕМ

4.1. Випадкові процеси. Ентропія як міра невизначеності інформації

Сигнал як випадковий процес

Сигнал – це матеріальний засіб перенесення інформації в просторі та часі. Сигнали в системах відіграють важливу роль. Вхідні енергетичні та матеріальні сигнали (потоки) живлять систему. Вхідні потоки інформації організовують роботу системи та здійснюють керування системою. Вихідні сигнали використовуються для вивчення структури та властивостей системи (ідентифікують систему).

Звичайна математична функція $f(t)$ не може служити моделлю сигналу $x(t)$. З означення функції відомо, що для кожного значення аргумента t однозначно відоме значення функції $f(t)$. Моделлю сигналу може бути лише набір функцій $f(t)$, причому до передачі сигналу невідомо, яку з них буде відправлено. Кожна конкретна функція часу називається ***реалізацією***. Така математична модель сигналу називається ***випадковим процесом*** (ВП). Одна окрема реалізація випадкового процесу називається ***випадковою величиною*** (ВВ).

Неперервні та дискретні в часі процеси. Випадковий процес із неперервним часом характеризується тим, що його реалізації можна визначити для всіх значень параметра t (*приклад* – температура). Дискретний в часі процес, задається на дискретному ряді точок часової осі (*приклад* – щоденний курс валют, щомісячна зарплата, щорічне значення врожайності). Дискретні у часі процеси записують у вигляді часових рядів.

Випадкові процеси можуть бути неперервними та дискретними також *за своїм значенням*.

Приклади: зміна температури – неперевний процес, оцінювання студента – дискретний процес.

Оскільки випадковий процес при фіксованих значеннях аргументу є випадковою величиною, то для його опису використовують такі самі способи, що й для випадкових величин: функції розподілу ймовірностей, щільності розподілу, моментні та кореляційні функції. Одновимірний закон розподілу достатньо повно характеризує випадковий процес у тому випадку, коли його значення при різних значеннях аргументу розглядаються ізольовано одне від одного (є незалежними). Найбільш часто на практиці зустрічається нормальний закон розподілу.

Стаціонарні процеси. Серед всіх випадкових процесів особливе значення мають стаціонарні процеси. Випадковий процес називається **стаціонарним** у широкому розумінні, якщо з плином часу залишаються незмінними його математичне сподівання та середньоквадратичне відхилення**.**

Ергодичні процеси. Випадковий процес називають **ергодичним,** якщо середнє значення, отримане усередненням однієї реалізації на тривалому проміжку часу співпадає із середнім значенням, отриманим усередненням всіх реалізацій на коротких часових інтервалах. Властивість ергодичності дуже важко довести і, зазвичай, її приймають як гіпотезу. *Приклади* процесів, які вважають ергодичними: підкидання монет, врожайність сільськогосподарської культур на різних ділянках одного регіону.

Ентропія як міра невизначеності інформації

Поведінка випадкового об’єкта характеризується невизначеністю. Для оцінки невизначеності введено кількісну міру, яку називають ***ентропією***. В системному аналізі ентропія H служить кількісною мірою безпорядку (різноманітності) в системі.

Розглянемо *приклад*. Нехай подія A може відбутися з імовірністю $p = 0{,}99$ і не відбутися з імовірністю $q = 0{,}01$, а інша подія B має ймовірності відповідно $p = 0{,}50$ і $q = 0{,}50$. Очевидно, що в першому випадку результат досліду «майже напевно» відомий (ентропія мала), а в другому невизначеність результату настільки велика (ентропія велика), що від прогнозування бажано взагалі відмовитися. Ентропія залежить як від кількості можливих станів, так і від ймовірності перебування в них. За міру невизначеності випадкового об'єкта A зі скінченною множиною можливих станів n беруть величину

$$H(A) = -\sum_{i=1}^{n} p_i \cdot log_2\, p_i\,, \tag{4.1}$$

яку називають ***ентропією*** випадкового об'єкта A з розподілом $\{p_i\}$ (Л. Больцман, 1877 рік).

У загальному випадку розглядають об'єкт A, який може перебувати у декількох станах $A_1, A_2, \ldots, A_n$ з відомими ймовірностями $p_1, p_2, \ldots, p_n$.

Приклад. В таблицях 4.1 і 4.2 наведено різні значення ймовірностей кліматичних станів поточного року.

Таблиця 4.1

	Дуже вологий	Вологий	Нормальний	Сухий
p_i	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$
$p_i\, log_2\, p_i$	$-\frac{3}{8}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{8}$

Таблиця 4.2

	Дуже вологий	Вологий	Нормальний	Сухий
p_i	$\frac{1}{32}$	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{32}$
$p_i\, log_2\, p_i$	$-\frac{5}{32}$	$-\frac{1}{4}$	$-0{,}17$	$-\frac{5}{32}$

У першому прикладі (табл. 4.1), згідно із співвідношенням (4.1), ентропія дорівнює

$$H(A) = -\sum_{i=1}^{4} p_i \cdot log_2\, p_i = \frac{3}{8} + \frac{1}{2} + \frac{1}{2} + \frac{3}{8} = 1{,}75\,.$$

У другому прикладі (табл. 4.2) маємо

$$H(A) = -\sum_{i=1}^{4} p_i \cdot log_2\, p_i = \frac{5}{32} + \frac{1}{4} + 0{,}17 + \frac{5}{32} \approx 0{,}73\,.$$

Як бачимо, у другому прикладі, невизначеність помітно менша. Це пояснюється тим, що ймовірність однієї з подій є близькою до 1 ($p_3 = \frac{7}{8}$). Максимальна невизначеність системи відповідає випадку, коли ймовірності всіх станів однакові і дорівнюють $p_i = \frac{1}{4}$. При цьому $H(A) = 2$.

Ентропія має наступні ***властивості***.

1. $H(p_1, p_2, \dots, p_n) = 0$ тоді й тільки тоді, коли ймовірність однієї з подій дорівнює одиниці (а інші ймовірності дорівнюють нулю). Це означає, що результат можна передбачити напевно, тобто немає жодної невизначеності. У всіх інших випадках ентропія додатна.

2. $H(p_1, p_2, \dots, p_n)$ набуває свого найбільшого значення у випадку, коли ймовірності всіх станів однакові (невизначеність максимальна) $p_1 = p_2 = \dots = p_n = \frac{1}{n}$.

Згідно другим законом термодинаміки кількість ентропії в ізольованій хаотичній системі може лише збільшуватися. Це відповідає зменшенню організованості такої системи. Максимальне значення ентропії відповідає рівноважному стану системи. Однак, процес переходу до рівноваги не є монотонним. У системах можуть відбуватися флуктуації. ***Флуктуація*** – це випадковий процес, в результаті якого організованість системи збільшується. Тому на окремих етапах ентропія системи може зменшитися, хоча загальна тенденція – збільшення ентропії. У відкритих системах ентропія може зменшуватися завдяки організуючим сигналам (впливам), які поступають в систему.

Принцип компенсації ентропії говорить про те, що ентропія відкритої системи може бути зменшена, якщо при цьому ентропія навколишнього середовища збільшиться. Отже безентропійних процесів не буває. Зменшення ентропії в одній частині системи приводить до її збільшення в іншій частині системи.

Приклад: людина зручно впорядковує свій спосіб життя, підвищує комфортність, але це приводить до збільшення кількості відходів і погіршення стану навколишнього середовища. Тому стійкий розвиток суспільства є неможливим доти, поки ми не обмежимо свій споживацький апетит, поки не зменшимо кількість відходів та навчимось жити у гармонії з природою.

Кожну керовану систему умовно можна розділити на управляючу підсистему і об’єкт управління. Для підвищення організованості (зменшення ентропії) у системі використовують деякі управляючі впливи. Результат впливу залежить від співвідношення різноманітності керованої системи і керуючої системи. Коли керівник приймає рішення, перед ним є множина варіантів прийняття рішень. Для вибору оптимального варіанту керівник повинен застосувати множину методів аналізу і оцінки цих варіантів. Уільям Росс Ешбі у 1959 р. довів теорему, яка відома як ***принцип необхідної різноманітності***. Для успішного вирішення задачі управління управляюча система повинна мати більшу різноманітність, ніж об’єкт управління. Це твердження є одним з основних принципів кібернетики.

4.2. Вимірювання кількості інформації

Розглянемо процес одержання інформації, який складається з двох етапів: передача сигналу і прийом сигналу. Нехай сигнал складається з послідовності окремих символів. До одержання чергового символу ситуація характеризується невизначеністю того, який символ буде відправлено. Ця невизначеність характеризується ***апріорною ентропією*** H_0. Після одержання символу y_k невизначеність зменшується. Якщо немає шуму (перешкод), вона взагалі зникає (оскільки точно відомо, що було передано символ $x_k = y_k$), а якщо шум є, ми не можемо бути впевненими, що отримали відправлений символ, і виникає невизначеність, що характеризується ***апостеріорною ентропією*** H_1.

Визначимо **кількість інформації** як міру знятої невизначеності: числове значення кількості інформації (I) про об'єкт дорівнює різниці його апріорної (H_0) і апостеріорної (H_1) ентропії, тобто

$$I = H_0 - H_1. \tag{4.2}$$

Приклад. Відомо, що кліматичний стан території може приймати 4 різних значення: сухий, нормальний, вологий, дуже вологий. Необхідно виконати прогноз кліматичного стану майбутнього року. Якщо відсутня допоміжна інформація, можна побудувати початковий прогноз у припущенні, що всі 4 стани клімату є однаково ймовірними $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. Апріорна ентропія становить $H\left(A\right) = 2$. Після отримання додаткової статистичної інформації, описаної у табл. 4.2 (ймовірність нормального року становить $\frac{7}{8}$, ймовірність вологого року становить $\frac{1}{16}$, ймовірності інших подій однакові і становлять $\frac{1}{32}$) ми можемо побудувати більш адекватний прогноз. Відповідна апостеріорна ентропія дорівнює $H(A) \approx 0{,}73$. Таким чином, кількість отриманої інформації становить

$$I = H_0 - H_1 = 2{,}00 - 0{,}73 = 1{,}27\ .$$

Одиниці вимірювання ентропії та кількості інформації

В наведених вище прикладах ми не вказували одиниць вимірювання ентропії. Однак, з означення кількості інформації I (4.2) випливає що інформація та ентропія мають однакові одиниці виміру. У дискретному випадку візьмемо за ***одиницю ентропії*** невизначеність випадкового об'єкта – такого, що

$$H(X) = -\sum_{i=1}^{n} p_i\, log_2\, p_i = 1\,. \qquad (4.3)$$

За одиницю ентропії (інформації) вибирають ентропію об'єкта з двома рівноймовірними станами (*наприклад,* підкидання монети). Цю одиницю називають **бітом**. Відповідні стани об'єкта позначають символами 0 та 1.

Більші одиниці інформації: 1 байт (8 бітів), 1 Кбайт (1024 байти), 1 Мбайт (1024 Кбайти), 1 Гбайт (1024 Мбайти), 1 Тбайт (1024 Гбайти) і т.д..

Розглянемо на прикладах застосування формули (4.3).

Приклад 1. Яка кількість інформації потрібна, щоб дізнатися результат підкидання монети? У даному випадку $n = 2$ і події рівноімовірні, тобто $p_1 = p_2 = 0{,}5$. Відповідно до (4.3): $I = -0{,}5 \cdot log_2\, 0{,}5 - 0{,}5 \cdot log_2\, 0{,}5 = 1$ біт.

Приклад 2. Гра «Відгадай оцінку». Викладач оцінив відповідь студента деякою оцінкою (2, 3, 4 або 5) і, не оголошуючи її, запропонував студентові відгадати цю оцінку. На питання викладач відповідає лише «так» або «ні». Яку кількість інформації студент повинен отримати, щоб взнати свою неоголошену оцінку? Як правильно побудувати процес відгадування?

Будемо вважати, що ймовірності всіх оцінок однакові. Оскільки $n = 4$, отже $p(A_i) = \dfrac{1}{4}$, $log_2\, p(A_i) = -2$ і $I = 2$ біт.

Таким чином, для повного зняття невизначеності щодо неоголошеної оцінки студенту необхідно 2 біти інформації.

Для відгадування оцінки необхідно використати принцип поділу пополам впорядкованої множини. Оптимальна процедура відгадування носить назву вибірковий каскад і складається з таких запитань:

1. «Оцінка більша за 3?»
2. Якщо отримана відповідь «Так», то наступне запитання: «Оцінка більша за 4?». Якщо у першому випадку отримана відповідь «Ні», то наступне запитання: «Оцінка більша за 2?».

Як бачимо, для відгадування оцінки достатньо 2-х, правильно сформульованих, запитань. Отже, кількість інформації (в бітах) чисельно дорівнює кількості питань з рівноймовірними бінарними відповідями, які необхідно задати, щоб повністю зняти невизначеність задачі.

Приклад 3. Випадковим чином виймається карта з колоди, в якій є 32 карти. Яка кількість інформації вимагається, щоб відгадати, яка це карта? Як побудувати процес відгадування? Для даної ситуації $n = 32 = 2^5$, значить, $k = 5$ і, отже, $I = 5$ біт. Послідовність запитань придумайте самостійно.

Закон Хартлі. Описана вище процедура дозволяє визначити кількість інформації в будь-якій подібній задачі. Якщо всі n результатів рівноімовірні, то $p_i = \frac{1}{n}$. Звідси слідує, що

$$I = \sum_{i=1}^{n} \frac{1}{n} log_2 n = log_2 n . \qquad (4.4)$$

Ця формула була виведена в 1928 р. американським інженером Р. Хартлі і носить його ім'я. Для виділення деякого елемента з множини, що містить n елементів, вимагається кількість інформації, яка дорівнює $log_2 n$. Частковим випадком застосування формули (4.4) є ситуація, коли $n = 2^k$; підставляючи це значення в (4.4) отримаємо: $I = k$ біт.

Саме ця ситуація була реалізована у прикладах 1–3. Аналізуючи результати розв'язку, можна прийти до висновку, що k дорівнює кількості питань з бінарними рівноймовірними відповідями, які визначали кількість інформації у задачах.

Міра Шеннона. К. Шеннон через 70 років після Л. Больцмана сформулював постулати теорії інформації і визначив міру інформації за формулою $I = -\sum_{i=1}^{n} p_i \cdot log_2 p_i$, де n – кількість станів системи; p_i – ймовірність (чи відносна частота) переходу системи в i-тий стан (сума всіх p_i дорівнює 1).

Лише згодом помітили, що ця формула інваріантна до формули інформаційної ентропії (4.1), що є підтвердженням системного зв'язку між цими фундаментальними поняттями.

Справедливість і достатня універсальність формул Хартлі та Шеннона підтверджується даними нейропсихології.

Приклад. Час t, за який піддослідний реагує на вибір одного предмета із n наявних, як виявилося, лінійно залежить від міри Хартлі: $t = 200 + 180 log_2 n$ (*мс*).

Приклад. Один із дослідів для визначення психофізіологічних реакцій людини полягав у тому, що перед піддослідним багато разів вмикали одну із n лампочок, на яку він мав вказати в ході експерименту. Виявилося, що середній час, необхідний для правильної відповіді, пропорційний не кількості n лампочок, а саме величині I, яка обчислюється за формулою Шеннона, де p_i – ймовірність ввімкнути лампочку номер i.

В загальному випадку, очевидно, що

$$I = -\sum_{i=1}^{n} p_i \cdot log_2 p_i \le log_2 n .$$

Інші підходи до вимірювання інформації. Багато науковців розглядають різноманітні кількісні міри для вимірювання змісту інформації:

- міра інформації, що базується на понятті мети (А. Харкевич та ін.);
- міра знань, що базується на понятті тезаурусу $T=<X, Y, Z>$, де X, Y, Z – множини відповідно імен, змісту і значень (прагматики) цих знань (Ю. Шрейдер та ін.);
- міра складності відтворення двійкових слів (А. Колмогоров та ін.);
- міра апостеріорних знань (Н. Вінер та ін.);
- міра успішності прийняття рішення (Н. Мойсеєв та ін.);
- міри інформаційної подібності та відмінності тощо.

Приклади. За міру (Колмогорова) відтворення двійкового слова y для заданого відображення f і заданих двійкових слів x із множини X можна взяти $H(f,y)=min|x|$, $x \in X$, $f(x)=y$. Тут $|x|$ – довжина двійкового слова x.

Якщо апріорі відомо, що деяка змінна належить інтервалу (0; 1), а апостеріорі, що вона належить інтервалу $(a,b) \subset (0;1)$, то мірою (Вінера) кількості інформації, видобутої з апостеріорного знання, може бути відношення міри (a,b) до міри (0; 1).

4.3. Спектральний аналіз та фільтрування сигналів

Сигнал $x(t)$ називають ***періодичним*** якщо він регулярно повторюється через час T

$$x(t)=x(t+T). \tag{4.5}$$

Значення T називають ***періодом сигналу***, обернену величину $f=\dfrac{1}{T}$ – ***частотою сигналу***. Строго періодичні сигнали зустрічаються природі і в техніці, але рідко зустрічаються в економіці. Найбільш характерним прикладом періодичного сигналу є ***гармонічний сигнал***, який описується співвідношенням

$$x(t) = a \cdot cos\left(\frac{2\pi t}{T}\right) + b \cdot sin\left(\frac{2\pi t}{T}\right) = A \cdot cos\left(\frac{2\pi t}{T} + \varphi_0\right), \quad (4.6)$$

де A – амплітуда сигналу (максимальне значення); φ_0 – початкова фаза.

Прикладами гармонічних сигналів є звук музичного інструмента, електромагнітна хвиля (теле- або радіосигнал, світло).

Більшість економічних процесів є циклічними але не періодичними (період не є постійним). *Приклади* циклічних процесів: курс валюти, курс цінних паперів, врожайність зернових культур. Такі процеси є послідовністю періодів росту і спадання. Але економічні цикли є нерегулярними. Амплітуда та період таких циклів змінюються випадковим чином.

Дослідження показали, що більшість циклічних процесів можна наближено представити у вигляді суми декількох гармонік з різними значеннями амплітуди і частоти. Для знаходження частот, які є найбільш характерними для отриманого сигналу, використовують ***спектральний аналіз*** сигналів. В його основі лежить математичний апарат під назвою ***гармонічний аналіз Фур'є***. Найчастіше технічні чи економічні сигнали є дискретними і мають вигляд часового ряду $\{x_i\} = x_1, x_2, \ldots, x_N$ (значення випадкової величини в послідовні моменти часу). Під гармонічним аналізом розуміють представлення кожного елемента сигналу у вигляді суми гармонічних функцій

$$x_i = a_0 + \sum_{k=1}^{n/2}\left(a_k\, cos\frac{2\pi k}{n} i + b_k\, sin\frac{2\pi k}{n} i\right) \quad i = 1, \ldots, n\,. \quad (4.7)$$

Співвідношення (4.7) називають ***дискретним перетворенням Фур'є*** (ДПФ). Тригонометричні функції, які входять у суму називають гармоніками. Тут $f = \dfrac{k}{n}$ – частота k-ої гармоніки. Основна частота (мінімальна) $f_1 = \dfrac{1}{n}$ відповідає максимальному періоду $T = n$. Всі інші частоти є кратними до неї. Максимальна частота відповідає мінімальному періоду ($T = 2$) і становить $f_{max} = \dfrac{1}{2}$. Отже, весь спектр частот знаходиться в інтервалі $f \in \left[{}^{1}\!/_{n};\quad {}^{1}\!/_{2}\right]$.

Гармонічний аналіз Фур'є створює образ сигналу у частотній області. Найчастіше цей образ подають у вигляді графіка залежності амплітуди гармонік від їх частоти (***спектрограма***), або залежності амплітуди від періоду (***періодограма***). Спектрограма дає уявлення про те, які частоти (періоди циклів) є найбільш характерними для даної системи. Спектрограма – це основний результат спектрального аналізу випадкового процесу.

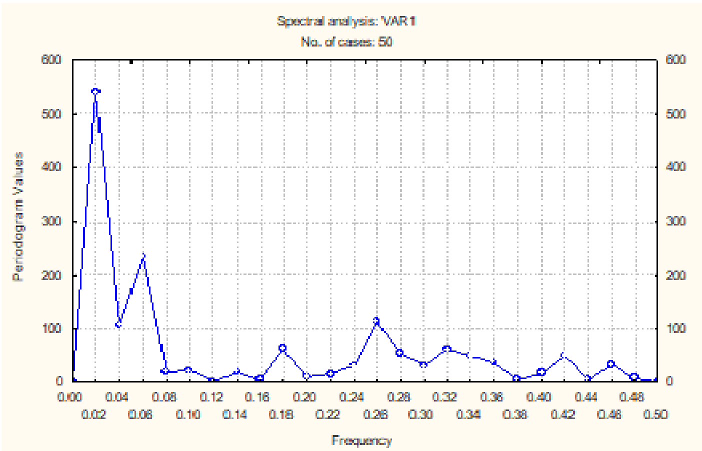


Рис. 4.1. Спектрограма ряду врожайності.
Основні частоти: 0.02; 0.06 і 0.25

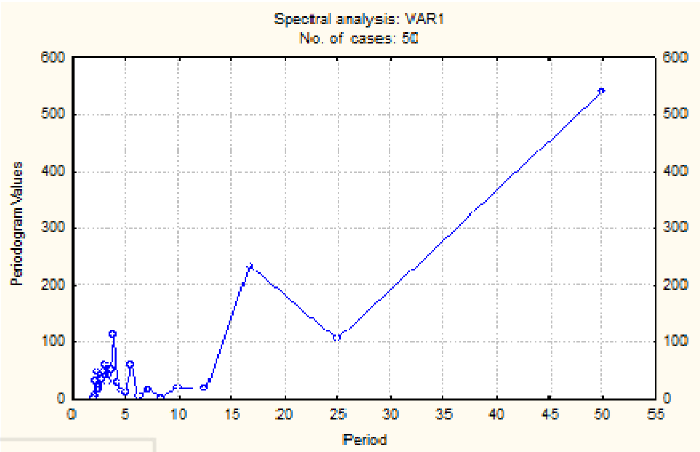


Рис. 4.2. Періодограма ряду врожайності. Основні періоди: 4; 16 і 50

Приклад. Розглянемо часовий ряд врожайності озимої пшениці в Україні ($n = 50$). Мінімальна частота $f_1 = \dfrac{1}{50} = 0{,}02$. Максимальна частота $f_{max} = \dfrac{1}{2} = 0{,}50$. Спектрограму ряду зображено на рис. 4.1, періодограму – на рис. 4.2.

Фільтрування сигналів

Для виділення переданого повідомлення із сигналу, спотвореного шумом, використовують фільтрування даних. Найчастіше використовується ***фільтр низьких частот***, який дозволяє очистити сигнал від високочастотних шумових спотворень. На вхід фільтра надходить сигнал, отриманий додаванням двох гармонік високої та низької частоти. Гармоніка низької частоти несе корисний сигнал, гармоніка високої частоти є шумом. Будуємо фільтр низьких частот у вигляді суми (4.7), верхня границя якої $k < K_{max}$. При цьому всі сигнали, частоти яких є вищими від $f_{\max} = \dfrac{K_{\max}}{T}$ не пропускаються фільтром (вилучаються з сигналу).

Прикладом фільтра є людське вухо. Воно пропускає лише частоти від 20 Гц до 20 Кгц. Слух інших живих істот дозволяє отримувати більш широкий спектр сигналів. Коти та собаки добре відчувають більш низькі звуки (інфразвуки), які виникають при землетрусах, ураганах і поширюються на великі відстані.

Фільтр низьких частот можна також використовувати для згладжування даних з метою виділення ***детермінованої складової*** часового ряду (*зауваження:* про часові ряди більш детально розповідається у наступному розділі). Часові ряди, які продукуються природними чи економічними системами, у значній мірі спотворюються стохастичними впливами (збуреннями). Для виявлення прихованої динаміки слід насамперед відфільтрувати високочастотні шуми, які є проявами стохастичної поведінки системи.

Найпростішим методом фільтрування високочастотних шумів є лінійне згладжування ряду з використанням прямокутного вікна заданої ширини w. Інша назва цього методу – ***метод ковзного середнього***. Його суть полягає в обчисленні для кожного елемента часового ряду його згладженого образу – середнього значення за сусідніми w даними (w – зазвичай, непарне число)

$$x_i^* = \frac{1}{w}\left(x_{i-v} + x_{i-v+1} + \ldots + x_i + \ldots + x_{i+v-1} + x_{i+v}\right),$$
$$v = \frac{w-1}{2}. \tag{4.8}$$

Чим більше значення w, тим більше даних беруть участь у розрахунку середнього, тим більш згладжена крива отримується (рис. 4.3). Для досягнення оптимального ефекту необхідно правильно підібрати ширину згладжуючого вікна.

Наприклад, при згладжуванні циклічних сигналів ширина згладжуючого вікна повинна становити половину періоду циклу.

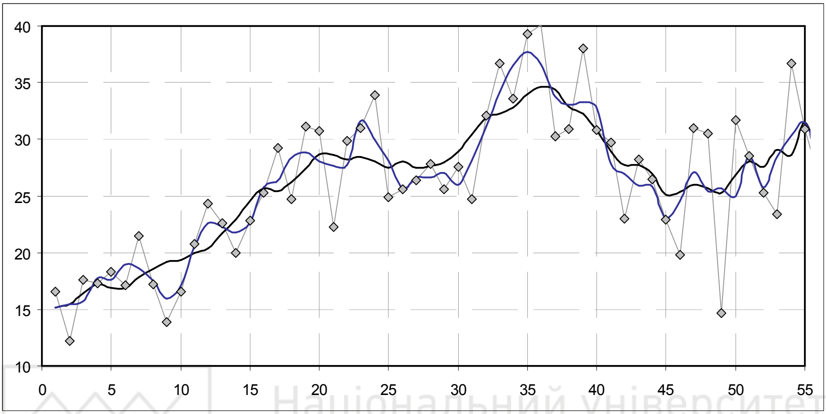


Рис. 4.3. Згладжування ряду врожайності методом ковзного середнього ($w = 3$; $w = 9$)

Вдале фільтрування сигналу полегшує його прогнозування. Математично доведено, що чим ширший спектр сигналу, тим швидше змінюється сигнал у часі, і навпаки, повільно змінні в часі сигнали мають вузький спектр. Звідси слідує висновок про те, що прогноз буде тим точнішим, чим вужчий спектр прогнозованого процесу. Загалом, будь-який прогноз – це неявне виділення низькочастотних складових досліджуваного процесу.

Приклад: згладжування часового ряду врожайності. Гармонічний аналіз показав, що для врожайності зернових в Україні є характерним цикл, тривалістю 18–20 років [7]. Але, недоліком методу ковзного середнього є те, що згладжений ряд має меншу довжину, ніж початковий. Це є суттєвим для рядів малої довжини. Щоб уникнути цього небажаного ефекту, початкові і кінцеві елементи ряду згладжуються на коротшому відрізку, ніж елементи середини ряду.

Наприклад, для ширини вікна $w = 9$ слід використовувати алгоритм згладжування, описаний співвідношеннями (4.9)

$$\begin{cases} y_1 = \dfrac{2 \cdot x_1 + x_2}{3};\ y_2 = \dfrac{x_1 + x_2 + x_3}{3};\ y_3 = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5}{5}; \\ y_4 = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7}{7}; \\ y_i = \dfrac{x_{i-4} + x_{i-3} + x_{i-2} + x_{i-1} + x_i + x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4}}{9}; \\ i = 5, \ldots, n-4; \\ y_{n-3} = \dfrac{x_{n-6} + x_{n-5} + x_{n-4} + x_{n-3} + x_{n-2} + x_{n-1} + x_n}{7}; \\ y_{n-2} = \dfrac{x_{n-4} + x_{n-3} + x_{n-2} + x_{n-1} + x_n}{5}; \\ y_{n-1} = \dfrac{x_{n-2} + x_{n-1} + x_n}{3};\ y_n = \dfrac{2 \cdot x_n + x_{n-1}}{3}, \end{cases} \tag{4.9}$$

де x_i – елементи вихідного ряду, y_i – елементи згладженого ряду, n – довжина ряду.

Іншим методом фільтрування часових рядів є ***метод експонентного згладжування***. Цей метод усуває недолік методу ковзного середнього, котрий полягає в тому, що всі дані, використовувані при обчисленні середнього, мають однакову вагу. Метод експонентного згладжування приписує найбільший ваговий коефіцієнт останньому спостереженню.

Визначимо величину α $(0 < \alpha < 1)$ як константу згладжування. Більше значення α приписує більшу вагу останнім спостереженням. Менше значення α посилює ефект згладжування ряду. Оптимальне значення параметра α можна знайти мінімізуючи похибку ретроспективного відтворення ряду. Нехай відомі значення часового ряду для минулих t моментів часу $y_1, y_2, \ldots, y_t$. Тоді оцінка y^*_{t+1} для моменту часу $t+1$ обчислюється за ітераційною формулою

$$y^*_{t+1} = \alpha y_t + (1-\alpha) y^*_t, \quad i = 3, 4, \ldots, n. \tag{4.10}$$

Стартове значення y_0^* знаходять шляхом усереднення декількох перших елементів часового ряду.

Теми для роздумів і контрольні запитання

1. Моделювання сигналу випадковим процесом.
2. Поняття ентропії як міри невизначеності. Властивості ентропії. Принцип компенсації ентропії.
3. В чому полягає принцип необхідної різноманітності Ешбі?
4. Яким чином вимірюється кількість інформації?
5. В яких одиницях вимірюється кількість інформації?
6. Закон Хартлі визначення кількості інформації.
7. Міра інформації Шеннона.
8. Який сигнал називають періодичним? Період і частота сигналу. Гармонічний сигнал.
9. Що таке спектральний аналіз? Спектрограма і періодограма випадкового процесу.
10. Поняття фільтрування сигналів.
11. Для чого використовують фільтри низьких частот?
12. Які Ви знаєте методи фільтрування сигналів?

5. АНАЛІЗ І ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

5.1. Часові ряди

Нехай відома статистика одного з параметрів системи за тривалий проміжок часу. Кожне значення параметра відповідає певному моменту часу, проміжок між моментами (часовий інтервал) – однаковий. Така таблиця називається часовим рядом.

Часовим рядом (**рядом динаміки**, англ. time series) називають послідовність значень статистичного показника (ознаки), впорядковану у хронологічному порядку. Окремі спостереження часового ряду, що відповідають певним моментам часу, називають його ***рівнями***, або ***елементами***. Рівні ряду можуть набувати як детермінованих, так і випадкових значень. Порядок розташування рівнів є істотною характеристикою ряду і не може змінюватися довільно.

Наприклад: ряд врожайності (часовий інтервал – 1 рік), ряд прибутків (часовий інтервал – 1 місяць), ряд температур (часовий інтервал – 1 день).

Іноді кожному моменту часу приводять у відповідність декілька значень різних показників досліджуваного об'єкта. Тоді отримують ***багатовимірний*** часовий ряд.

Кожен часовий ряд містить важливу інформацію про систему, на основі досліджень якої його було сформовано.

Наприклад, часовий ряд прибутків деякої фірми відображає вплив таких чинників системи, як інвестиції, кількість транспортних одиниць, рівень менеджменту фірми тощо.

Досліджуючи часовий ряд, ми отримуємо інформацію про такі характеристики системи як її розмірність, характер динаміки (хаотична, циклічна) та ін. Аналіз часових рядів, як правило, передбачає проведення таких основних етапів [7; 30]:

- графічне подання і попередній аналіз поведінки часового ряду;
- виокремлення і видалення закономірних складових ряду (тренду, сезонних та циклічних компонент);
- виокремлення і видалення низько- та високочастотних складових (фільтрація);

- дослідження випадкової складової часового ряду;
- побудова загальної моделі досліджуваного ряду;
- дослідження отриманої моделі і прогнозування майбутньої поведінки системи, що вивчається;
- вивчення взаємодії між різними часовими рядами, що характеризують систему чи процес.

При дослідженні часових рядів застосовують наступні методи.

1. ***Кореляційний аналіз***, який дає змогу виявляти істотні періодичні залежності та їх лаги (затримки) всередині певного процесу (автокореляція) або між декількома процесами (кроскореляція).

2. ***Спектральний аналіз***, що застосовують для визначення періодичних та квазіперіодичних компонент часового ряду.

3. ***Методи згладжування та фільтрації***, що призначені для перетворення часових рядів з метою видалення з них високочастотних та сезонних коливань.

4. ***Методи авторегресії та ковзних середніх***, які використовують для опису і прогнозування процесів, що здійснюють випадкові коливання навколо певного середнього значення.

5. ***Методи прогнозування***, що дають можливість на основі обраної моделі часового ряду оцінювати його найбільш імовірні значення в майбутньому.

Існує два підходи до прогнозування часових рядів. Перший – **метод регресії**. В цьому випадку будуються формули виду

$$y = f\left(x_1, x_2, \ldots, x_n\right), \tag{5.1}$$

які відображають залежність досліджуваної величини y від інших параметрів x_i ($i = \overline{1,n}$) у явному вигляді. Підставляючи у формулу (5.1) очікувані у майбутньому значення параметрів x_i, які мають суттєвий вплив на систему, отримуємо прогноз розвитку системи на майбутній період.

При іншому підході розглядається лише залежність ключового показника y від часу t. Звичайно, час не є чинником даного процесу, змінна часу t просто акумулює вплив всіх умов і причин, які визначають цей процес. Дослідження показали, що передісторія часового ряду відображає впливи головних чинників і обумовлює можливість моделювання та прогнозування поведінки часових рядів без врахування цих чинників. Такий підхід називається **аналізом динаміки часового ряду**.

Залежно від характеру часового параметра розрізняють ***моментні*** та ***інтервальні*** часові ряди. У моментних рядах рівні характеризують значення показника станом на певний момент часу. Їм відповідають змінні типу запасу (англ. stock variables).

Наприклад, моментними є наступні ряди: ціна товару, курс валюти, кількість населення регіону, маса об’єкта, електрична напруга у мережі, температура хворого тощо.

В інтервальних рядах рівні характеризують значення показника за певні періоди (інтервали) часу. Їм відповідають змінні типу потоку (англ. *flow variables*).

Прикладами є ряди динаміки виробництва продукції (річний, квартальний, місячний, добовий), міграції населення, прибутку підприємства, валового національного доходу, розсіюваної потужності, витрат реагентів у хімічному реакторі тощо.

Часовий ряд, сформований з перших різниць рівнів моментного ряду, є інтервальним. Часовий ряд, рівні якого отримані підсумовуванням наростаючим підсумком всіх рівнів інтервального ряду, починаючи з певного визначеного рівня, є моментним. Співвідношення між моментними та інтервальними часовими рядами є аналогічним до співвідношення між функціями та їх похідними.

Рівні рядів динаміки можуть бути ***абсолютними, відносними або середніми значеннями*** певних показників. Якщо вони не є величинами, вимірюваними безпосередньо, а похідними від них – середніми чи відносними, то відповідні ряди також називають похідними.

Наприклад, похідними є ряди середньодобового або середньомісячного виробництва певної продукції.

Особливістю інтервальних рядів є можливість підсумовування їх рівнів. Результатом цього є так звані нагромаджені підсумки, які є значеннями певного показника за відповідний збільшений період.

Наприклад, підсумовуючи добові обсяги виробництва продукції протягом місяця, отримаємо загальний обсяг її виробництва за відповідний місяць.

Суми рівнів моментного ряду не мають змісту. Проте різниця рівнів моментного ряду є зміною показника за відповідний період.

Приклад. Якщо рівні ряду є значеннями середньодобової температури, то, підсумовуючи їх протягом певного періоду, ми не отримаємо величини, яка б мала фізичний зміст. Для цього нам потрібно поділити результат на кількість діб у досліджуваному періоді. Тоді ми отримаємо величину, що є середнім значенням температури за відповідний період.

Успішність аналізу часового ряду істотно залежить від правильності вибору інтервалу між його сусідніми рівнями. Зручніше вибирати рівновіддалені рівні ряду, але це не завжди є можливим.

Наприклад, офіційні курси валют і фондові індекси не визначаються у вихідні та святкові дні, тому рівні відповідних рядів не є рівновіддаленими.

Вибір надмірно великого інтервалу може призвести до втрати інформації про суттєві особливості ряду. Крім того, кількість членів ряду при цьому може виявитися недостатньою для застосування деяких методів аналізу. З іншого боку, надмірно малі інтервали збільшують обсяг розрахунків і можуть затіняти особливості динаміки показника випадковими флуктуаціями.

Важливою умовою аналізу є ***порівнянність*** рівнів ряду. Їх непорівнянність може бути зумовлена змінами законодавства та нормативної бази, методик розрахунку показників, меж адміністративно-територіальних одиниць, технологій виробництва, термінологій, природними, екологічними та техногенними чинниками. Інформація про ряд має бути достатньо повною. Зокрема, для аналізу рядів, що містять циклічну складову, залежно від застосовуваних методів математичної обробки, потрібна інформація щодо проміжку, який перевищує 3–6 повних циклів. При побудові регресійних моделей необхідно мати ряди, довжина яких у кілька разів перевищує кількість параметрів, що визначаються.

Рівні часових рядів можуть включати аномальні значення (викиди). Причинами їх появи найчастіше є помилки під час збору та обробки інформації. Але викиди також можуть бути зумовлені реальними стрибками в динаміці досліджуваних показників [30].

При аналізі та моделюванні часових рядів найчастіше використовують **адитивну модель** часового ряду

$$y_t = v_t + s_t + c_t + \varepsilon_t \,. \tag{5.2}$$

Складовими цієї моделі є параметри, що визначають основні характеристики процесу.

Параметр v_t – характеризує тренд. **Трендом** називають стійку систематичну зміну процесу протягом тривалого часу. Тренд відображає вплив чинників довготривалої дії.

Наприклад: за відсутності інвестицій прибуток транспорної фірми з часом почне зменшуватись через зношеність автомобілів.

Оцінка тренду може здійснюватися параметричними і непараметричними методами. ***Параметричний тренд*** полягає у підборі простої функції (лінійної, квадратичної, логарифмічної), яка б описувала тенденцію зміни ряду. ***Непараметричний*** метод використовують, коли важко підібрати відповідну функцію. Тоді виконують згладжування ряду методом ковзного середнього або іншим.

Параметр s_t – це **сезонна складова**. Вона відображає регулярну повторюваність процесів у часі (як правило, за період, що не перевищує року).

Наприклад, залежність кількості пасажирів від пори року.

Параметр c_t – це **циклічна складова**, що описує тривалі періоди відносного спаду або підйому.

Наприклад, зміна врожайності під впливом багаторічних циклів сонячної активності (11 років).

Для моделювання циклічної та сезонної складових, як правило, використовують гармонічний аналіз.

Параметр ε_t – це **випадкове відхилення** (**білий шум**), що пояснюється дією багатьох випадкових, важко передбачуваних чинників.

В результаті аналізу часового ряду необхідно визначити, які з вказаних вище компонент присутні у ньому та побудувати для них ефективні моделі. Моделі детермінованих компонент v_t, s_t, c_t, зазвичай, є аналітичними (деякі функції). Моделлю випадкового відхилення ε_t є відповідний закон розподілу випадкової величини.

Циклічна компонента часового ряду

Наявність періодичної чи сезонної складової ряду означає, що кожне чергове спостереження x_i є схожим на спостереження x_{i-L}, яке відмічалось деякий час (рік, місяць, 10 років) тому. Така залежність може бути виявлена як кореляційна залежність порядку L між кожним i-им елементом ряду і його $i-L$-им елементом. Таку залежність називають **автокореляцією**, тобто кореляцією між членами одного і того ж ряду. L називають **лагом** (лаг – зсув, запізнення). Циклічні складові часового ряду можуть бути визначені за допомогою дослідження автокореляційної функції.

Автокореляційна функція (АКФ) – це послідовність коефіцієнтів автокореляції, які відповідають значенням лагу від 1 до k. Графічним зображенням автокореляційної функції є ***корелограма*** (залежність $\rho(L)$, $L=\overline{1,k}$). Як правило послідовність лагів обмежують $1/4$ довжини початкового ряду

($k \leq N/4$). Для побудови корелограми в MS Excel використовують функцію **КОРРЕЛ (***масив 1;масив 2***)**.

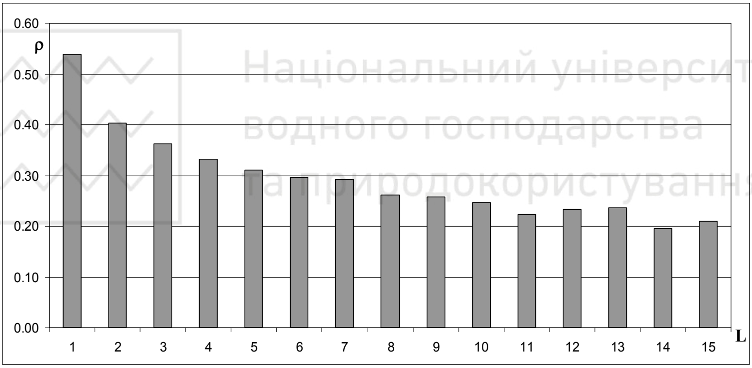
В залежності від автокореляційної функції розрізняють такі види часових рядів:

1) якщо АКФ містить багато значущих елементів, говорять про ряд з довгою пам'яттю (рис. 5.1, а); характерним прикладом є ряд ціни товару;

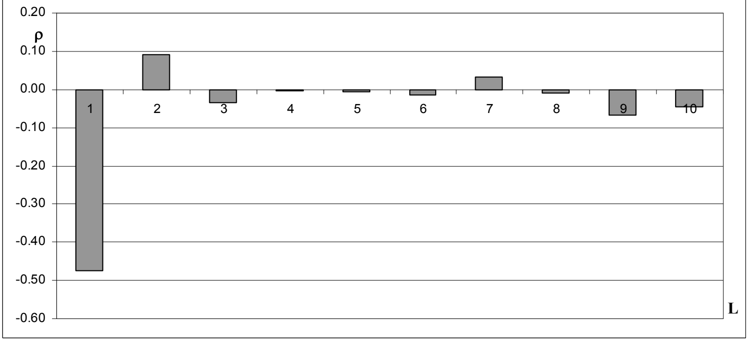
2) якщо АКФ містить 1–2 значущих елементи, говорять про ряд з короткою пам'яттю (рис. 5.1, б);

3) якщо більшість значень АКФ є незначущими, але один з них є великим позитивним – маємо наявність циклу (рис. 5.1, в);

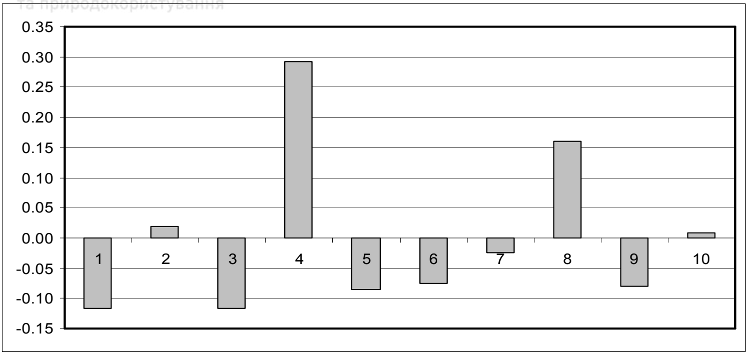
4) якщо більшість значень АКФ є незначущими, то це свідчить про незалежність елементів часового ряду.



а)



б)



в)

Рис. 5.1. Корелограми часових рядів
а) ряд з довгою пам'яттю (розливи річки Ніл);
б) ряд з короткою пам'яттю (зміни врожайності озимої пшениці, США);
в) циклічний ряд (річні суми опадів, Рівненська область)

Для моделювання циклічних компонент часового ряду застосовують гармонічний аналіз Фур'є. При цьому припускають, що заданий часовий ряд відповідає одному коливанню $(T = N)$ і кожен елемент часового ряду x_i представляють у вигляді

$$x_i^* = a_0 + \sum_{k=1}^{T/2}\left(a_k \cos\frac{2\pi k}{T} i + b_k \sin\frac{2\pi k}{T} i\right), \quad i = 1, \ldots, T\,. \quad (5.3)$$

Весь спектр частот знаходиться в інтервалі $f \in \left[1/N ; 1/2 \right]$.

Для комп'ютерного моделюваня більш зручним є практичний гармонічний аналіз. При цьому кількість гармонік в залежності (5.3) невелика: 3–4. Період коливань T кожної гармоніки визначають методом повного перебору з деяким кроком ΔT, коефіцієнти a_k, b_k визначають за методом найменших квадратів. Значення періоду T, яке найбільш точно відображає коливання досліджуваного ряду, вибирають з мінімального значення похибки $\sum_{i=1}^{N}\left(x_i - x_i^*\right)^2$ (рис. 5.2).

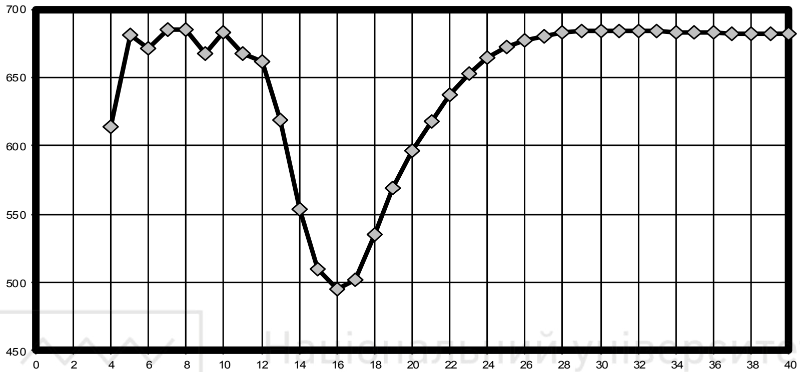


Рис. 5.2. Залежність похибки гармонічної моделі від періоду гармоніки

Стаціонарні часові ряди

Якщо математичне сподівання і дисперсія часового ряду не змінюються з часом, такий ряд називають **стаціонарним** (у широкому розумінні). Стаціонарний ряд можна розглядати як набір випадкових величин і вивчати його методами математичної статистики. На практиці стаціонарні ряди зустрічаються дуже рідко. Будь-який економічний чи природний процес, як правило, містить деяку тенденцію розвитку, якій відповідає тренд часового ряду. Шляхом вилучення тренду та сезонної і циклічної складових часовий ряд можна перетворити в стаціонарний.

Формально цей процес виглядає так.

1. Для початкового ряду y_n моделюємо тренд t_n.

2. Вилучаємо тренд. Отримуємо ряд первинних залишків $E_{1n} = y_n - t_n$.

3. Досліджуємо ряд залишків E_{1n} на стаціонарність. Якщо ряд не є стаціонарним переходимо до дослідження циклічності.

4. Моделюємо циклічну складову.

5. Вилучаємо циклічну складову. Отримуємо ряд вторинних залишків $E_{2n} = E_{1n} - c_n$.

6. Досліджуємо якість побудованої моделі $y_n^* = t_n + c_n$.

Трендостійкі, реверсивні та випадкові часові ряди

Ряд може складатися із трендостійких ділянок та ділянок з випадковою (стохастичною) динамікою. Для дослідження типу динаміки ряду досліджують фактичну частоту коливань (коливність) ряду. Для цього використовують метод поворотних точок.

Поворотною точкою називають значення ряду x_t для якого виконується одна з двох умов:

$$x_{t-1} < x_t; \quad x_{t+1} < x_t \text{ або } x_{t-1} > x_t; \quad x_{t+1} > x_t.$$

М. Кендал показав, що для випадкового (стохастичного) ряду довжиною n середня кількість поворотних точок становить $P_C = 2 \cdot \dfrac{n-2}{3}$ при стандартному відхиленні

$$\delta = \sqrt{\frac{16n-29}{90}}.$$

Розрізняють такі типи рядів за Кендалом.

1. Якщо $P < P_L$, де $P_L = P_C - 2\delta$, то це означає, що для ряду характерні трендові ділянки. Такий ряд називають ***персистентним*** (***трендостійким***).

2. Якщо $P > P_U$, де $P_U = P_C + 2\delta$, то це означає, що велика кількість коливань викликана деяким зовнішнім чинником. Такий ряд називають ***реверсивним***.

3. Якщо $P_L < P < P_U$ – ряд вважають ***випадковим*** (***стохастичним***).

5.2. Методика прогнозування стану системи

Упродовж усієї історії людства багато мислителів і провидців намагалися «зазирнути» у майбутнє, щоб мати уявлення про його можливий подальший розвиток. Для цього використовували найрізноманітніші методи – від хіромантії і чаклунства до астрологічних та науково-фантастичних передбачень. Блискучі приклади передбачення майбутнього демонстрував у своїй діяльності знаменитий Леонардо да Вінчі. Його геніальний розум породив фантастичні ідеї підводних човнів, літаків, велосипедів, ріжучих верстатів і навіть танків, які через кілька століть стали не тільки можливими, але й дуже наближеними до реальних.

У зв’язку з викликами і загрозами, що все частіше постають перед сучасним суспільством, виникла потреба на об’єктивних засадах передбачати хоча б приблизні сценарії майбутніх подій. Це потрібно для формування раціональної і безпомилкової (чи менш помилкової) стратегії розвитку фірми, галузі, чи цілої країни [13, С. 427–428].

Для прийняття правильних управлінських рішень необхідно оцінити розвиток системи на майбутнє. Побудована оцінка повинна адекватно відповідати майбутнім подіям.

Наприклад, в управлінні перевезеннями необхідно спрогнозувати рівень прибутку в майбутньому на основі поточних даних чи очікувану кількість пасажирів на майбутній місяць тощо.

Сучасні спроби одержати об’єктивні знання про майбутнє пов’язані переважно з розробленням нових і застосуванням традиційних методів математики і статистики. До них належить ціла група потужних методів, як-от метод часових рядів, методи регресійного аналізу, імітаційного моделювання тощо. Усі вони належать до класу так званих *методів кількісного прогнозування*, і їх застосовують для приблизного «визначення» майбутньої поведінки певної змінної величини або системи взаємопов’язаних змінних величин на заздалегідь відомому часовому інтервалі.

Незважаючи на найширшу практику використання цих методів і наявність величезних обчислювальних потужностей, їхнє практичне застосування принципово обмежено лише випадками обробки ретроспективних даних кількісного характеру про монотонно змінні процеси. Інакше кажучи, у разі застосування методів прогнозування відбувається описання майбутнього, яке фактично є продовженням або екстраполяцією минулого [13, С. 428–429].

Прогноз – науково обґрунтоване передбачення ймовірності можливих станів системи в майбутньому, шляхів та термінів її розвитку. Очевидно, що майбутнє неможливо спостерігати, а очікуваний результат – виміряти, його можна лише передбачити за певних умов, скажімо, «*…якщо тенденція не зміниться, то…*» або «*…якщо станеться подія* A*, то…*» і т. ін. Якщо умови зміняться, то автоматично зміниться й результат прогнозування. Отже, статистичний прогноз, побудований за схемою «*…якщо, то…*», завжди є умовним.

Іншою особливістю статистичного прогнозу є визначеність його в часі. Часовий горизонт прогнозу називають ***періодом упередження***. За тривалістю цього періоду вирізняють прогнози:

- **оперативний** (короткотерміновий) – розрахований на час, протягом якого не очікується суттєвих кількісних та якісних змін системи (термін прогнозування – від 1 місяця до 1 року);
- **довгостроковий** – розрахований на час, протягом якого очікуються істотні кількісні і якісні зміни системи (термін – від 5 до 20 років);
- **середньостроковий** – охоплює час між оперативним та довгостроковим прогнозом (до 5 років) з переважанням кількісних змін над якісними.

Прогнозний результат на період упередження можна представити одним числом (***точковий прогноз***) або інтервалом значень, до якого з певною ймовірністю належить прогнозна величина (***інтервальний прогноз***).

Статистичні прогнози ґрунтуються на гіпотезах про стабільність значень величини, що прогнозується; незмінність закону її розподілу; стабільність взаємозв'язків з іншими величинами тощо. Основний інструмент прогнозування –

екстраполяція. Суть прогнозної екстраполяції полягає в поширенні закономірностей, зв'язків і відношень, виявлених у спостереженому періоді, за його межі.

Залежно від гіпотез щодо механізму формування і подальшого розвитку процесу використовуються різні методи прогнозної екстраполяції. Їх можна об'єднати в дві групи.

♦ **Екстраполяція закономірностей розвитку** — тенденцій і коливань. Екстраполяція закономірностей розвитку ґрунтується на вивченні передісторії, виявленні усталених тенденцій, траєкторій зміни в часі. Абстрагуючись від причин формування процесу, закономірності його розвитку розглядають як функцію часу. Інформаційною базою прогнозування слугують одновимірні часові ряди.

♦ **Багатофакторне прогнозування**. Процес розглядається як функція певної множини факторів (чинників). Інформаційною базою виступає система взаємозв'язаних часових рядів. При цьому особливого значення набуває теоретичний аналіз структури взаємозв'язків параметрів.

Важливим етапом статистичного прогнозування є ***верифікація прогнозів***, тобто оцінювання їх точності та обґрунтованості.

Найбільш поширене *ретроспективне оцінювання прогнозу*, тобто оцінювання прогнозу для минулого часу. Часовий ряд поділяється на дві частини: перша – для $t = 1, 2, 3, ..., p$ – називається ретроспекцією (передісторією), друга (прогнозний період) – для $t = p+1, p+2, p+3, ..., p+(n-p)$. За даними ретроспекції моделюється закономірність динаміки і на основі моделі розраховується прогноз $x^*_{p+1}, x^*_{p+2}, \ldots, x^*_{p+v}$, де v — період упередження. Оскільки фактичні значення прогнозного періоду відомі, то можна визначити похибку прогнозу як різницю фактичного x_t і прогнозного x_t^* рівнів: $e_t = x_t - x_t^*$.

Узагальнюючою оцінкою точності прогнозу є середня відносна похибка MAPE (Mean Absolute Percentage Error)

$$\bar{e} = \frac{\sum |e_t|}{n - p}. \tag{5.4}$$

Якщо результат оцінювання точності прогнозу задовольняє визначені критерії точності, скажімо, 10%, то прогнозна модель вважається прийнятною і рекомендується для практичного використання.

При прогнозуванні процесів, розвиток яких не піддається формалізації (наприклад, розвиток науки і техніки, соціально-економічні та політичні наслідки прийняття певних управлінських рішень), використовують методи експертних оцінок. Вони ґрунтуються на професійному досвіді експертів, які добираються за принципом компетентності.

5.3. Прогнозування методами ковзного середнього, експонентного згладжування, екстраполяції тренду

Прогнозування часових рядів методом ковзного середнього. Досить поширеним і простим методом аналізу динаміки є згладжування ряду. Суть його полягає в заміні фактичних рівнів x_t середніми за певними інтервалами. При цьому характер динаміки проявляється чіткіше. Процедуру згладжування називають фільтруванням. На практиці використовують переважно лінійні фільтри, з-поміж яких найпростіший – ***метод ковзного середнього*** з інтервалом згладжування $m < n$. У цьому методі кожний член ряду замінюється середнім значенням m сусідніх членів, де m – ширина «вікна». Згладжений ряд можна розглядати як тренд, змодельований механічним чином. Інтервали поступово зміщуються на один елемент: $x_1, x_2, ..., x_m$; $x_2, x_3, ..., x_{m+1}$; $x_3, x_4, ..., x_{m+2}$ тощо. Для кожного з них визначається середня $\bar{x}_t$, яка припадає на середину інтервалу. Графік згладженого ряду має вигляд (рис. 5.3).

Якщо m – непарне число, тобто $m = 2p + 1$ то

$$x_t^* = \frac{1}{2p+1}\sum_{t-p}^{t+p} x_i , \tag{5.5}$$

де x_i – фактичне значення рівня в i-й момент; i – порядковий номер рівня в інтервалі.

Наприклад, при ширині вікна 5 кожне значення ряду y_t замінюється згладженим значенням

$$x_t^* = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5} . \tag{5.6}$$

У симетричних фільтрах старіша і новіша інформація рівнозначимі, проте, при прогнозуванні важливішою є більш нова інформація. У такому разі використовують асиметричні фільтри. Найпростіший з них – ковзна середня, яка замінює не центральний, а останній член ряду (*адаптивна середня*). При цьому формула (5.5) має наступний вигляд (ширина вікна 5)

$$x_{t+1}^* = \frac{x_{t-4} + x_{t-3} + x_{t-2} + x_{t-1} + x_t}{5} . \tag{5.7}$$

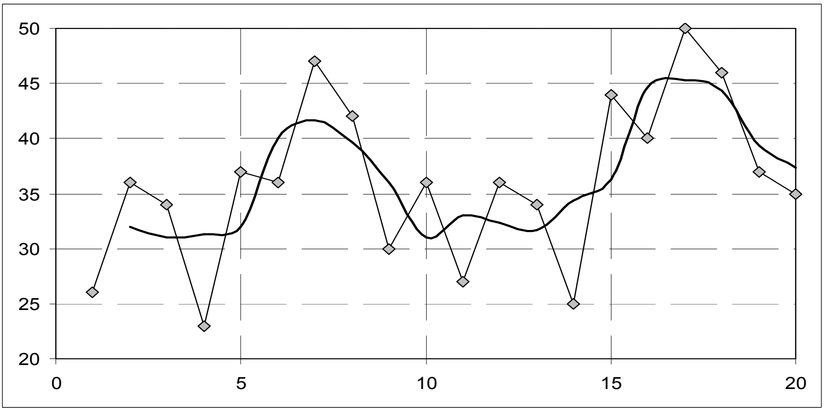


Рис. 5.3. Моделювання часового ряду методом ковзного середнього (база 3 роки)

Не існує чіткого правила для вибору числа m – бази методу, що використовує ковзне середнє. Якщо графічний аналіз вказує на слабкий тренд, вибирають більші значення m. У разі сильного тренду вибирають менші значення m. Якщо ряд має циклічну складову, ширина згладжуючого вікна m повинна дорівнювати половині циклу.

Приклад. Розглянемо часовий ряд прибутків фірми (тис. грн.) за 21 місяць (табл. 5.1).

Таблиця 5.1

Місяць	Прибутки	Місяць	Прибутки	Місяць	Прибутки
1	15	8	28	15	42
2	20	9	30	16	37
3	17	10	38	17	41
4	22	11	27	18	38
5	25	12	39	19	36
6	30	13	30	20	41
7	24	14	35	21	37

Якщо використати значення $m = 3$ як базу ковзного середнього, то оцінка прибутку на наступний місяць ($t = 21$) буде дорівнювати середній величині прибутку за 18, 19 і 20 місяці

$$y^*_{21} = \frac{38 + 36 + 41}{3} = 38{,}33 \ .$$

Оцінкою якості прогнозної моделі є її **точність**, яка оцінюється похибкою прогнозування. Використаємо відоме значення прибутку за 21 місяць для оцінки похибки прогнозування. Визначаємо абсолютну та відносну похибку прогнозу.

Абсолютна похибка – це різниця між прогнозом і фактичним значенням параметра $\Delta = \left|38{,}33 - 37\right| = 1{,}33$.

Відносна похибка дорівнює відношенню абсолютної похибки прогнозу до фактичного значення параметра. Відносна похибка отриманого нами прогнозу дорівнює $\delta = \dfrac{38{,}33 - 37}{37} = 0{,}036$ або 3,6%. Зауважимо, що, оскільки, прогнозування є випадковим процесом, оцінку точності моделі не можна робити за результатом одного прогнозу.

Прогнозування методом експонентного згладжування.

У методі ковзного середнього всі елементи часового ряду вважаються однаково важливими. Насправді, зазвичай, найважливішими є останні значення параметра, оскільки саме на них опирається прогноз. Цю особливість даних враховує метод експонентного згладжування. Даний метод приписує спостереженням вагові коефіцієнти, причому найбільший коефіцієнт присвоюється останньому спостереженню.

Визначимо величину α $(0 < \alpha < 1)$ як константу згладжування. Більше значення α означає більшу вагу останніх спостережень. Менше значення α означає посилення ефекту згладжування.

Нехай відомі значення часового ряду для минулих t моментів часу $x_1, x_2, ..., x_t$. Тоді оцінка x^*_{t+1} для моменту часу $t+1$ обчислюється за ітераційною формулою

$$x^*_{t+1} = \alpha x_t + \left(1-\alpha\right) x^*_t, \quad i = 3, 4, ..., n. \tag{5.8}$$

Початковий елемент згладженого ряду x^*_0 вибирається шляхом усереднення декількох перших елементів часового ряду.

Для попереднього прикладу (табл. 5.1) при $\alpha = 0{,}2$ оцінка для $t = 21$ дорівнює

$$x^*_{21} = \alpha x_{20} + \left(1-\alpha\right) x^*_{20} = 0{,}41 \cdot 41 + 0{,}59 \cdot 37{,}39 = 38{,}91 \; .$$

Відносна похибка прогнозу становить 5,2%.

Графік часового ряду, змодельованого методом експонентного згладжування, зображено на рис. 5.4.

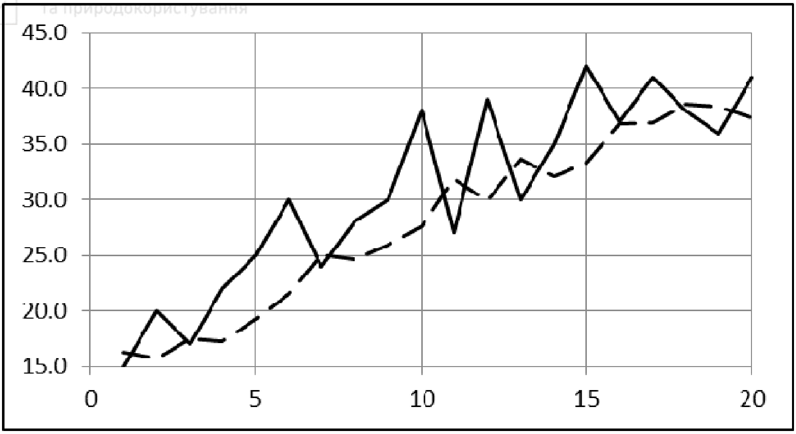


Рис. 5.4. Моделювання часового ряду методом експонентного згладжування ($\alpha = 0,41$). Штрихова лінія – згладжений ряд

Прогнозування методом екстраполяції тренду

Розглянемо аналітичні методи моделювання тренду. Регресійний аналіз дозволяє встановити аналітичний зв'язок між зміною якоїсь величини (наприклад, прибутку) і часом. Найпростіша регресійна модель описує лінійну залежність прибутку від часу

$$x^* = a_1 t + a_0 . \tag{5.9}$$

Константи a_1 і a_0 визначаються з часового ряду за ***методом найменших квадратів***.

Щоб перевірити, наскільки лінійна модель відповідає вихідним даним, необхідно обчислити ***коефіцієнт кореляції*** r ($0 \le |r| \le 1$). Чим ближчим є $|r|$ до 1, тим кращою є лінійна модель.

Крім лінійної моделі тренду використовують і нелінійні моделі: поліноміальну (найчастіше квадратичну), логарифмічну, експоненціальну. Для вибору оптимальної трендової моделі використовують низку критеріїв.

1. Критерій R^2. Для коефіцієнта детермінації R^2 має місце: $0 \le R^2 \le 1$. Більше значення коефіцієнта детермінації означає кращу відповідність моделі до вихідних даних.

2. Критерій найменших квадратів. Критерієм виступає сума квадратів відхилень модельних значень від фактичних SSE (Sums of Squares Errors). Менше значення суми означає кращу якість моделі

$$S = \sum_{t=1}^{n} \left(x_t - x_t * \right)^2 . \qquad (5.10)$$

3. Критерій відносної похибки. Критерієм виступає середнє значення відносної похибки побудованої моделі MAPE (Mean Absolute Percentage Error). Менше значення похибки означає кращу якість моделі

$$\bar{\varepsilon} = \frac{\sum_{t=1}^{n} \left| x_t - x_t^* \right|}{x_t} . \qquad (5.11)$$

Отриману трендову модель можна використовувати для прогнозування методом екстраполяції.

Покажемо це на прикладі лінійної регресії. Лінійна модель прибутку для наведених вище даних представляється формулою

$$x^* = 1{,}256 \cdot t + 17{,}558 . \qquad (5.12)$$

При $t = 21$ одержуємо $x^* = 1{,}256 \cdot 21 + 17{,}558 = 43{,}93$. Отже прогноз прибутку на 21-й місяць становить 43,93 тис. грн.

Коефіцієнт кореляції лінійної моделі $r = 0{,}89$, коефіцієнт детермінації $R^2 = 0{,}79$. Високе значення коефіцієнта детермінації вказує на хорошу якість моделі. Сума квадратів відхилень SSE моделі становить $S = 276{,}04$. Середня відносна похибка MAPE моделі $\overline{\varepsilon} = 0{,}101$.

Розглянемо модель квадратичного тренду (рис. 5.5). Рівняння квадратичного тренду має вигляд

$$x^{*} = -0{,}057 \cdot t^{2} + 2{,}456 \cdot t + 13{,}159\,. \qquad (5.13)$$

Коефіцієнт детермінації моделі $R^{2} = 0{,}835$. Сума квадратів відхилень $S = 218{,}74$. Середня відносна похибка моделі $\overline{\varepsilon} = 0{,}090$.

Отже, за всіма показниками квадратична модель є кращою від лінійної. Слід, однак, зауважити, що поліноміальні моделі рідко використовують для прогнозування. Причиною є сильні осциляції поліноміальних кривих.

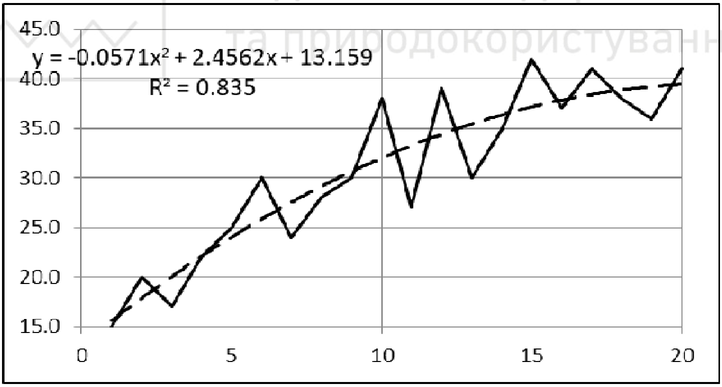


Рис. 5.5. Квадратична модель тренду часового ряду

Отримане вище значення x^{*} є точковим прогнозом. Перевагою трендової регресійної моделі у порівнянні з іншими є можливість побудови інтервального прогнозу. При цьому з надійністю у p % гарантують, що прогнозована величина x не вийде з інтервалу $x^* - \Delta x \le x \le x^* + \Delta x$. Для знаходження ширини довірчого інтервалу використовують формулу

$$\Delta x = t_\alpha \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(n+1-\bar{t})^2}{\sum (t-\bar{t})^2}}\ , \qquad (5.14)$$

де $S = \sqrt{\dfrac{\sum\limits_{t=1}^{n}\left(x_t - x_t^*\right)^2}{n-2}}$; $x_t^* = a_1 t + a_0$; $t_c = \dfrac{1}{n}\sum t$ (при p = 95% маємо $t_\alpha = 2.10$).

У нашому випадку $S = 3{,}92$; $t_{5\%} = 2{,}10$;. $\Delta x = 2{,}10 \cdot 3{,}92 \cdot 0{,}46 = 3{,}82$. Отже шуканий довірчий інтервал лінійної прогнозної моделі має вигляд: $43{,}93 \pm 3{,}82$ або $x \in [40{,}11;\ 47{,}75]$. Зауважимо, що для побудови довірчого інтервалу слід спочатку перевірити нормальний закон розподілу ряду залишків.

5.4. Оцінка якості моделі

Оцінка якості трендової моделі виконується шляхом аналізу ряду залишків. Залишок трендової моделі розрахуємо за формулою

$$e_t = x_t - x_t^*. \qquad (5.15)$$

Модель є **адекватною**, якщо ряд залишків має властивості ***випадковості***, ***незалежності послідовних рівнів***, ***нормальності розподілу*** і ***рівності нулю середнього значення***.

Проілюструємо алгоритм оцінки якості моделі на наведеному вище прикладі.

1. Оцінка *близькості залишку до нуля* виконується на основі t***-критерію*** значущості (критерію Ст'юдента). Якщо виконується умова $t < t_{кр}$, то середнє значення можна вважати статистично близьким до нуля. Тут $t = \dfrac{\bar{e}\sqrt{n}}{\sigma_e}$, а $t_{кр}$ можна визначити за допомогою функції Excel **СТЬЮДРАСПОБР**$(\alpha, n-1)$.

У нашому прикладі $\bar{e} = 0{,}017$, $\sigma_e = 3{,}39$, $t = 0{,}022$, $t_{кр} = 2{,}10$. Оскільки виконується умова $t < t_{кр}$, то середнє значення залишку можна вважати статистично близьким до нуля.

2. Перевірку *випадковості* рівнів ряду залишків проведемо на основі ***критерію поворотних точок***. Кількість поворотних точок у досліджуваному ряду $P = 12$. У випадковому ряду повинна виконуватися подвійна нерівність:

$$P_L < P < P_U. \tag{5.16}$$

При $n = 20$: $P_L = 10$, $P_U = 14$. Отже, має місце $10 < P < 14$ і властивість ***випадковості*** виконується.

3. Перевірка *незалежності* рівнів здійснюється за допомогою ***d-критерію Дарбіна–Уотсона***

$$d = \frac{\sum_{t=2}^{n}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{N} e_t^2}. \tag{5.17}$$

У нашому прикладі $d = 3{,}07$. При $n = 20$ критичні табличні величини $d_L = 1{,}20$ і $d_U = 1{,}41$; $4 - d_U = 2{,}59$ і $4 - d_L = 2{,}80$. Оскільки $d > 4 - d_L$, присутня негативна автокореляція залишків. Отже, властивість ***незалежності залишків*** не підтверджена.

4. Перевіримо відповідність ряду залишків e_t нормальному закону розподілу. Це можна зробити за допомогою RS ***-критерію***

$$RS = \frac{e_{max} - e_{min}}{S}, \qquad (5.18)$$

де e_{max} – максимальне значення ряду залишків; e_{min} – мінімальне значення ряду залишків; S – середньоквадратичне відхилення значень ряду залишків.

Якщо розраховане значення RS потрапляє в інтервал між верхньою та нижньою межею із заданим рівнем ймовірності, то гіпотеза про нормальний розподіл приймається.

Для нашого випадку $RS_{\min} = 3,18$; $RS_{\max} = 4,49$; $RS = 3,61$. Оскільки $RS \in [RS_{min}; RS_{max}]$, то *критерій відповідності* ряду залишків e_t до ***нормального закону розподілу*** виконується.

З чотирьох перевірених нами критеріїв виконуються три. Не виконується лише критерій незалежності залишків. Отже, модель часового ряду можна покращити, якщо врахувати зв'язок, який існує між послідовними елементами ряду прибутків. Для цього можна використати авторегресійну модель або гармонічну модель.

Теми для роздумів і контрольні запитання

1. Що називають часовим рядом? Рівні ряду.
2. Порядок аналізу часових рядів.
3. Які методи застосовують при дослідженні часових рядів?
4. Для чого використовують кореляційний та спектральний аналіз часових рядів?
5. Для чого призначені методи згладжування та фільтрації?
6. Методи авторегресії та ковзних середніх використовують для…
7. Які підходи існують до прогнозування часових рядів?

8. Чим відрізняються моментні та інтервальні часові ряди? Наведіть приклади.
9. Адитивна модель часового ряду. Які параметри є її складовими? Які методи використовуються для оцінки параметрів адитивної моделі?
10. Поняття автокореляції. Автокореляційна функція. Ряд з довгою пам’яттю, ряд з короткою пам’яттю, циклічний ряд.
11. Моделювання циклічної компоненти часового ряду.
12. Стаціонарні часові ряди. Як (теоретично) можна перетворити часовий ряд в стаціонарний?
13. Трендостійкі, реверсивні та випадкові часові ряди за Кендалом. Поняття поворотної точки ряду.
14. Чим відрізняються точковий та інтервальний прогнози? Наведіть приклади.
15. В чому суть прогнозної екстраполяції?
16. Яка відмінність між екстраполяцією закономірностей розвитку і багатофакторним прогнозуванням?
17. Що таке верифікація прогнозу і які методи для цього застосовують?
18. Прогнозування часових рядів методом ковзного середнього.
19. Прогнозування часових рядів методом експонентного згладжування.
20. Прогнозування часових рядів методом екстраполяції тренду.
21. Оцінка якості моделі часового ряду.

6. СТАТИСТИЧНІ МЕТОДИ ОБРОБКИ ДАНИХ

6.1. Випадкова величина та її числові характеристики

На практиці часто доводиться мати справу з величинами, які набувають певних, заздалегідь не передбачуваних, значень. Такі величини називають ***випадковими***. Якщо випадкова величина має скінченну кількість можливих значень $x_1, x_2, ..., x_m$ то її називають дискретною випадковою величиною. В іншому разі маємо неперервну випадкову величину.

Сукупність усіх можливих значень дискретної випадкової величини називається ***генеральною сукупністю***. Генеральну сукупність обробляти важко, або ж неможливо, в силу її великого обсягу. Тому на практиці генеральну сукупність заміняють меншим обсягом випадково вибраних з неї даних. Результати обмеженого ряду спостережень $x_1, x_2, ..., x_n$ ($n << m$) випадкової величини називають ***вибіркою*** з генеральної сукупності.

Вибірка повинна об'єктивно представляти генеральну сукупність, тобто бути ***репрезентативною***. Для цього:

а) вона повинна бути випадковою;

б) вона повинна представляти всі частини генеральної сукупності.

Найважливішою характеристикою випадкової величини є ***закон розподілу***, який встановлює відповідність між кожним можливим значенням дискретної випадкової величини $x_1, x_2, ..., x_m$ та частотою її появи $p_1, p_2, ..., p_m$. Графічно емпіричний закон розподілу можна представити у вигляді ***гістограми частот***.

Для кількісної характеристики закону розподілу неперервної випадкової величини вводять поняття ***функції розподілу*** $F(x)$ як ймовірності того, що випадкова величина X набуває значень, менших від деякого числа x

$$F(x) = p(X < x). \tag{6.1}$$

Іншою характеристикою неперервної випадкової величини X є ***щільність розподілу*** ймовірності її появи

$$f(x)=F'(x). \tag{6.2}$$

Найбільш поширеним законом розподілу неперервних випадкових величин є ***нормальний закон розподілу***. Графічним зображенням його щільності розподілу є симетрична дзвоноподібна крива (крива Ґаусса) (рис. 6.1).

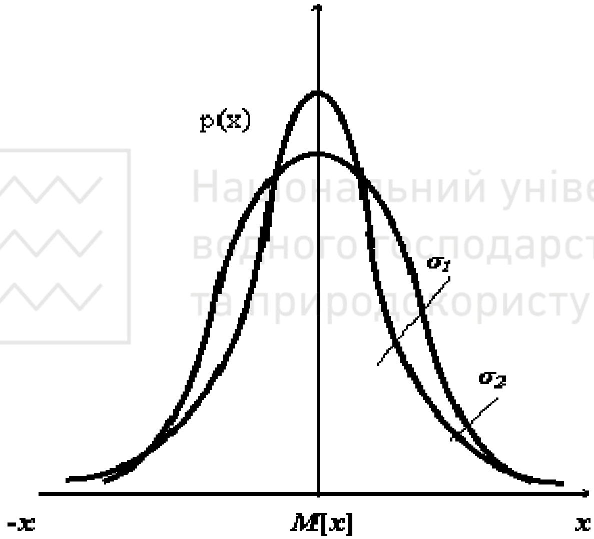


Рис. 6.1. Щільність розподілу випадкової величини з нормальним законом розподілу

Основними числовими характеристиками нормального закону розподілу є математичне сподівання $M(x)$, дисперсія σ_x^2 та середнє квадратичне відхилення σ_x. Математичним сподіванням дискретної випадкової величини називається сума добутків усіх можливих її значень на їх ймовірності

$$M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i .$$

Математичне сподівання відповідає значенню випадкової величини, яке спостерігається найчастіше. На практиці його замінюють середнім статистичним значенням вибірки $\overline{x}$. ***Середнє квадратичне відхилення*** характеризує міру розсіювання випадкової величини навколо середнього значення. Для нормального розподілу справедливе правило «3-х сігм»: в інтервал $\left[\overline{x} - 3\sigma_x ; \overline{x} + 3\sigma_x\right]$ попадає $99{,}7\%$ всіх значень випадкової величини. $95{,}4\%$ значень випадкової величини x зосереджено в інтервалі $\left[\overline{x} - 2\sigma_x < x < \overline{x} + 2\sigma_x\right]$. $68{,}3\%$ значень випадкової величини x знаходяться в інтервалі $\left[\overline{x} - \sigma_x < x < \overline{x} + \sigma_x\right]$.

Основні числові характеристики вибірки розраховують наступним чином.

Середнє вибіркове значення випадкової величини x визначаємо за формулою

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i . \quad (6.3)$$

Дисперсія σ_x^2 характеризує розсіювання випадкової величини навколо середнього значення $\overline{x}$

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 . \quad (6.4)$$

При малих об’ємах вибірки x $(n \le 30)$ використовують виправлену дисперсію

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 .$$

Середнє квадратичне відхилення σ_x є більш зручною характеристикою розсіювання випадкової величини, оскільки його розмірність співпадає з розмірністю випадкової величини

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} . \qquad (6.5)$$

Кореляційний зв'язок випадкових величин

Однією з найголовніших задач системного аналізу є встановлення взаємозв'язків між різними параметрами, які описують поведінку системи. *Наприклад*, необхідно перевірити, чи збільшення кількості коштів, виділених на техогляд та ремонт автомобілів, сприятиме збільшенню доходу фірми, що надає послуги перевезення. Для того, щоб ефективно управляти такими зв'язками, треба вміти вимірювати ці зв'язки кількісно.

У процесі вивчення систем часто доводиться на підставі великої кількості досліджуваних даних виявляти найбільш суттєві чинники, які впливають на досліджуваний об'єкт, а також встановлювати форму зв'язку між різними, зв'язаними один з одним чинниками. Для спрощення такого аналізу виділяють один суттєвий параметр системи і вивчають його залежність від інших, *наприклад*, залежність прибутку фірми від коштів, виділених на ремонт, від кількості автомобілів тощо. Оскільки досліджувані величини є випадковими, то така залежність буде нестійкою (кореляційною).

У найпростішому випадку досліджують залежність деякої випадкової величини Y від найбільш важливого чинника X. Нехай в результаті статистичних досліджень отримали таблицю залежності (табл. 6.1).

Таблиця 6.1

Значення випадкової величини Y
за відповідних значень чинника X

X	x_1	x_2	x_3	$\ldots$	x_i	$\ldots$	x_n
Y	y_1	y_2	y_3	$\ldots$	y_i	$\ldots$	y_n

Аналізуючи задану таблицю необхідно відповісти на два запитання:

1) чи є зв'язок між випадковими величинами X і Y;

2) який цей зв'язок.

Тобто, треба знайти аналітичний вигляд функції $y = f(x)$, яка правдоподібно описувала б цю таблицю дослідних даних.

Першим кроком дослідження зв'язку є візуальний графічний аналіз. На координатній площині $X0Y$ будуються точки з координатами (x_i, y_i). Отримуємо точковий графік або поле розсіювання величин X і Y. Характер розташування точок може дати наочне уявлення про тип (вигляд) формули, яка зв'язує досліджувані величини і тісноту зв'язку між ними.

У *прикладі*, поданому на рис. 6.2 у першому випадку (а) маємо лінійний зв'язок, у другому (б) – нелінійний (квадратичний), у третьому (в) – можна говорити, що зв'язок відсутній, оскільки точки рівномірно розсіялися по координатній площині.

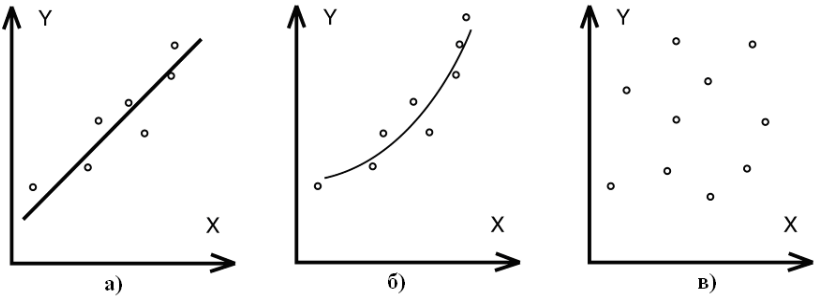


Рис. 6.2. Поле розсіювання величин X і Y:
а) лінійний зв'зок; б) квадратичний зв'зок; в) зв'язок відсутній

Найбільш тісним зв'язком є ***функціональний*** (всі точки лягають на одну лінію – графік відповідної функції). За наявності функціонального зв'язку кожному значенню параметра X відповідає лише одне значення параметра Y. Це означає що на чинник Y найбільший вплив має саме чинник X, а вплив інших чинників малосуттєвий. Приклади таких залежностей можна часто зустріти у техніці, *наприклад*, зв'язок між масою конструкції та частотою власних коливань.

Проте, в економічних та соціальних системах функціональний зв'язок є рідкістю, що пояснюється складністю цих систем. Набагато частіше в таких системах зустрічається неоднозначний, або ***кореляційний*** зв'язок (*наприклад*, зв'язок між прибутком Y та кількістю виділених коштів X). При кореляційній залежності зміна однієї величини викликає зміну середнього значення іншої. Але при цьому, одному і тому ж значенню чинника X можуть відповідати різні значення чинника Y. Це пояснюється неврахованим впливом інших чинників. Тісноту кореляційного зв'язку між двома випадковими величинами x та y характеризує **коефіцієнт коваріації**

$$\operatorname{cov}(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \qquad (6.6)$$

6.2. Лінійна регресія. Верифікація моделі лінійної регресії

Встановити вигляд формули $y = f(x)$ залежності між двома випадковими величинами x та y дозволяє ***регресійний аналіз***. Формулу $y = f(x)$ називають емпіричною (взятою з досліду), або ж рівнянням регресії чинника Y на чинник X. Для побудови рівняння регресії найчастіше підбирають найпростіші функції: лінійну, квадратичну, логарифмічну тощо. Поняття регресії є одним із ключових понять статистичного аналізу даних.

Під **регресією** розуміють односторонню стохастичну залежність однієї випадкової величини від іншої або ж від декількох випадкових величин.

Емпіричні регресійні формули мають велике практичне значення. Вдало підібрана емпірична формула дає змогу вирішувати задачу прогнозування, яке є основою для прийняття управлінських рішень.

Процес побудови емпіричних формул складається з двох етапів:

1) встановлення загального виду формули регресії;
2) визначення найкращих параметрів регресійної формули.

Вибір функції візуально підказує поле розсіювання. Значення коефіцієнтів емпіричної формули оцінюють за методом найменших квадратів (МНК).

Принцип найменших квадратів. Сума квадратів відхилень емпіричної функції $y = f(x_i)$ від експериментальних значень y_i повинна бути мінімальною

$$S = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \to min . \tag{6.7}$$

Найбільш поширеними статистичними моделями є лінійні. Розглянемо випадок лінійної залежності між X і Y

$$y^* = a_1 x + a_0 , \tag{6.8}$$

де коефіцієнти a_1 та a_0 – невідомі.

Згідно з принципом найменших квадратів

$$S(a_0, a_1) = \sum_{i=1}^{n} \left(y_i - (a_1 x_i + a_0)\right)^2 \to \min . \tag{6.9}$$

Умовою мінімуму функції $S(a_0, a_1)$ є

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = 0 .$$

Звідси отримуємо систему двох лінійних алгебричних рівнянь відносно невідомих a_1 та a_0

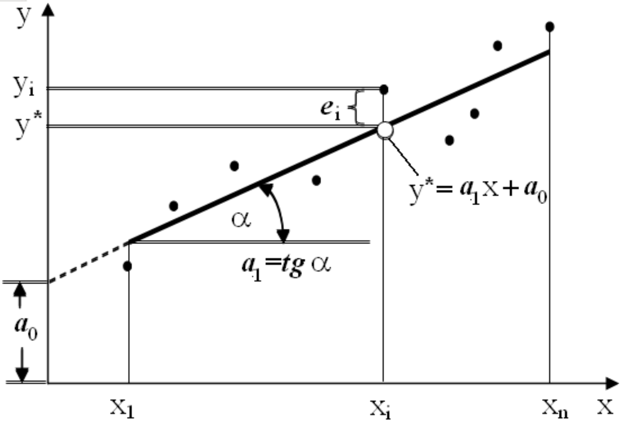
$$\begin{cases} a_1 \sum\limits_{i=1}^{n} x_i^2 + a_0 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ a_1 \sum\limits_{i=1}^{n} x_i + a_0 n = \sum\limits_{i=1}^{n} y_i. \end{cases} \tag{6.10}$$

Розв'язавши систему (6.10), знаходимо невідомі коефіцієнти

$$a_1 = \frac{\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i - n \sum\limits_{i=1}^{n} x_i y_i}{\left(\sum\limits_{i=1}^{n} x_i \right)^2 - n \sum\limits_{i=1}^{n} x_i^2};$$

$$a_0 = \frac{\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i^2 \sum\limits_{i=1}^{n} y_i}{\left(\sum\limits_{i=1}^{n} x_i \right)^2 - n \sum\limits_{i=1}^{n} x_i^2}. \tag{6.11}$$

Графіком рівняння регресії (6.8) є пряма регресії (рис. 6.3).



x	y
x_1	y_1
x_2	y_2
---	---
x_i	y_i
---	---
---	---
x_n	y_n

Рис. 6.3. Графік рівняння лінійної регресії

Параметр a_0 називається ***зсувом (зміщенням)***, а a_1 – ***нахилом***. Математична інтерпретація цих параметрів зрозуміла з наведеного рисунка.

Передумови застосування методу найменших квадратів.

Важливу роль при оцінці якості регресійної моделі відіграють залишки моделі

$$e_i = y_i - y_i^*, \tag{6.12}$$

тобто невраховані моделлю y^* впливи інших чинників. Для застосування МНК є необхідним виконання декількох умов.

1. Математичне сподівання залишків моделі дорівнює нулю.
2. Значення ряду залишків незалежні між собою і мають постійну дисперсію.
3. Пояснювальні змінні моделі не пов’язані із залишками.

4. Пояснювальні змінні $x_1, x_2, ..., x_m$ є незалежними.

Верифікація моделі лінійної регресії

Емпірична модель ***парної лінійної регресії*** має наступний вигляд

$$y = a_0 + a_1 x + e\,, \tag{6.13}$$

де $y = (y_1, y_2, ..., y_n)$ – ряд спостережень за досліджуваним параметром; $x = (x_1, x_2, ..., x_n)$ – ряд спостережень за впливаючим параметром (чинником або фактором); $e = (e_1, e_2, ..., e_n)$ – ряд залишків (6.12), у якому $y_i^* = a_1 x_i + a_0$ – значення параметра, розраховані за рівнянням лінійної регресії.

Отриману модель регресії необхідно перевірити на адекватність і статистичну значущість, тобто виконати верифікацію моделі з метою обґрунтування її якості і можливості її подальшого використання.

Спочатку визначають ***коефіцієнт парної кореляції*** r_{yx} за формулою

$$r_{yx} = \frac{cov(x, y)}{\sigma_x \sigma_y}\,, \tag{6.14}$$

де σ_x, σ_y – середні квадратичні відхилення змінних x і y

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}\,, \sigma_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}\,.$$

Коефіцієнт кореляції є *критерієм щільності лінійного взаємозв'язку* між змінними x і y. Значення коефіцієнта кореляції розташовані в діапазоні між «–1» та «+1» ($-1 \le r_{yx} \le 1$). Додатне значення коефіцієнта кореляції свідчить про прямий зв'язок між змінними (із зростанням однієї випадкової величини зростає середнє значення іншої), а від'ємне – про зворотний зв'язок. Коли коефіцієнт кореляції прямує за модулем до 1, це свідчить про наявність сильного лінійного зв'язку між x і y; у протилежному випадку, коли коефіцієнт кореляції прямує до 0, лінійного зв'язку немає.

Для перевірки *статистичної значущості коефіцієнта кореляції* за ***t–критерієм Стьюдента*** необхідно розрахувати t–відношення

$$t_r = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \tag{6.15}$$

та порівняти його з критичним значенням параметра $t_{kp} = t_{\alpha/2}, n-2$, де α – рівень значущості, зв'язаний з рівнем надійності P співвідношенням $\alpha = 1 - P$. Якщо $|t_r| > t_{kp}$ – коефіцієнт r_{xy} є значущим (тобто, статистично значуще відрізняється від нуля).

Ще однією характеристикою лінійного регресійного зв'язку є ***коефіцієнт детермінації*** R^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \tag{6.16}$$

Тут $SST = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$ – загальна сума квадратів; $SSE = \sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 = \sum_{i=1}^{n} e_i^2$ – сума квадратів похибок моделі;

$SSR = \sum_{i=1}^{n} \left(y_i^* - \bar{y} \right)^2$ – сума квадратів, що пояснюється рівнянням регресії.

Між коефіцієнтом детермінації R^2 та коефіцієнтом парної кореляції r_{yx} є проста залежність

$$R^2 = r_{yx}^2 . \qquad (6.17)$$

Коефіцієнт детермінації є *критерієм адекватності моделі* до статистичних даних і показує, на скільки відсотків варіація залежної змінної y визначається варіацією незалежної змінної x.

Наприклад, значення коефіцієнта детермінації $R^2 = 0{,}95$ свідчить про те, що 95% загальної дисперсії змінної y пояснюється рівнянням регресії, а 5% – іншими неврахованими в моделі випадковими чинниками, що свідчить про високу адекватність побудованої моделі до статистичних даних.

Перевірка адекватності моделі здійснюється також на основі F *–**критерію Фішера***. Спочатку розраховується F – статистика

$$F = \frac{R^2}{1 - R^2} \left(n - 2 \right). \qquad (6.18)$$

Потім для рівня значущості α (як правило, він дорівнює 0,05 або 0,01) та ступенів свободи $k_1 = 1; \, k_2 = n - 2$ визначається критичне значення критерію Фішера $F_{kp} = F_{\alpha, 1, n-2}$.

Якщо виконується співвідношення $F > F_{kp}$, то побудована регресійна модель адекватно апроксимує дані спостережень.

Необхідно також здійснити перевірку статистичної значущості параметрів моделі a_0 та a_1. Для кожного з коефіцієнтів необхідно розрахувати t–статистику

$$t_{a_0}^* = \frac{a_0}{S_{a_0}}, \quad t_{a_1}^* = \frac{a_1}{S_{a_1}}. \tag{6.19}$$

Тут S_{a_0}, S_{a_1} – середні квадратичні відхилення параметрів a_0 та a_1

$$S_{a_1} = \frac{S}{\sqrt{n \cdot \sigma_x}}, S_{a_0} = S_{a_1}\sqrt{\frac{\sum x_i^2}{n}}\ , \tag{6.20}$$

де параметр $S = \sqrt{\dfrac{\sum(y_i - y_i*)^2}{n-2}}$ – середнє квадратичне відхилення залишків регресійної моделі.

При рівні значущості α для ступеня свободи $k = n-2$ визначається критичне значення t–розподілу Ст'юдента $t_{кр} = t_\alpha$. Якщо виконується умова $\left|t_{a_j}^*\right| > t_{кр}$, то відповідний параметр вважається статистично значущим. У протилежному випадку параметр є статистично незначущим і повинен бути вилученим з моделі. Якщо статистично незначущим є параметр a_1, то неадекватною є модель парної лінійної регресії у цілому.

Для статистично значущих параметрів моделі можна побудувати ***інтервали довіри*** – інтервали, в які із заданою ймовірністю p (рівнем довіри) попадають оцінені нами значення параметрів моделі a_0 і a_1. Ці інтервали визначаються за наступними формулами

$$a_0 - t_\alpha \cdot S_{a_0} < \alpha_0 < a_0 + t_\alpha \cdot S_{a_0},$$
$$a_1 - t_\alpha \cdot S_{a_1} < \alpha_1 < a_1 + t_\alpha \cdot S_{a_1}. \qquad (6.21)$$

Прогнозування на основі моделі лінійної регресії

Якщо доведена адекватність і статистична значущість моделі парної лінійної регресії, то її можна використовувати для прогнозування значення залежної змінної y. При цьому можна отримати два види прогнозів: ***точкові*** та ***інтервальні***.

Точковий прогноз для будь-якого значення чинника x_{n+1} дає середнє очікуване значення змінної y

$$\overset{*}{y}_{n+1} = a_0 + a_1 \cdot x_{n+1}. \qquad (6.22)$$

Ймовірність реалізації точкового прогнозу дуже мала. Тому, зазвичай, використовують інтервальний прогноз, який дає можливість визначити інтервал, у який з ймовірністю p можуть потрапити майбутні значення змінної y. Ширина довірчого інтервалу Δy визначається співвідношенням

$$\Delta y = t_{\alpha/2} \cdot S \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}}. \qquad (6.23)$$

Нижня і верхня межа інтервального прогнозу визначаються таким чином

$$y_{min} = \overset{*}{y}_{n+1} - \Delta y; \quad y_{max} = \overset{*}{y}_{n+1} + \Delta y. \qquad (6.24)$$

Однією з важливих економічних характеристик досліджуваного процесу є ***коефіцієнт еластичності***, який для лінійної регресії обчислюється за наступною залежністю

$$E = a_1 \frac{\overline{x}}{\overline{y}}. \qquad (6.25)$$

Коефіцієнт еластичності показує на скільки відсотків зміниться досліджуваний параметр y при зміні чинника x на 1%.

Для полегшення розрахунків параметрів регресії можна використати вбудовану функцію табличного редактора MS Excel **ЛИНЕЙН()**. Це матрична функція, результати якої представлені у вигляді таблиці розміром 5 рядків і $m+1$ стовпців, де m – кількість параметрів моделі. Для оцінки лінійної регресії в Excel-таблиці необхідно виділити 5 рядків та 2 стовпці і виконати:

Мастер функцій / ***Статистические*** / ***ЛИНЕЙН()***.

Завершуємо розрахунок одночасним натисканням комбінації клавіш CTRL+SHIFT+ENTER.

Приклад результату, отриманого за допомогою функції **ЛИНЕЙН(),** наведено в табл. 6.2.

Критерій Дарбіна–Уотсона

При побудові моделі парної лінійної регресії можливі порушення деяких передумов класичного лінійного регресійного аналізу. Одним з таких можливих порушень є ***автокореляція залишків*** – взаємозв'язок між послідовними значеннями залишків. Автокореляція залишків приводить до неефективності МНК–оцінок параметрів та неефективності прогнозів (прогнози з великою дисперсією).

Оскільки автокореляція є негативним явищем, потрібно вміти її тестувати. Найбільш поширеним тестом перевірки моделі на наявність автокореляції залишків є тест Дарбіна-Уотсона. Цей тест має наступний алгоритм.

1. На основі МНК будується економетрична модель і знаходиться ряд залишків e_i, $\left(i = \overline{1,n}\right)$.

Таблиця 6.2

Розрахунок і оцінка лінійної регресії з використанням функції MS Excel **ЛИНЕЙН()**

	t_{a_1}	t_{a_0}					
	9,39	4,04					
	a_1	a_0					
$a =$	**0,505**	**271,1**	Коефіцієнти регресії				
$m_a =$	0,054	67,044	Похибки коефіцієнтів регресії				
$R^2 =$	**0,831**	39,733	Коефіцієнт детермінації і стандартна помилка оцінки y				
$F =$	**88,25**	18	Критерій Фішера і кількість ступенів свободи				
$\sigma =$	139323	28417	Пояснена дисперсія і дисперсія залишків				
$t_{кр} =$	**2,10**		Критичне значення критерію Ст'юдента				
$F_{кр} =$	**4,41**		Критичне значення критерію Фішера				

2. Розраховується критерій Дарбіна–Уотсона DW

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}. \tag{6.26}$$

3. Задаючи рівень значущості α, для моделі з m чинниками і кількістю спостережень n за таблицями розподілу Дарбіна–Уотсона визначаються два значення d_L і d_U.

4. Будують зони автокореляційного зв'язку, які схематично можна представити у наступному вигляді (рис. 6.4)

Рис. 6.4. Схема зон автокореляційного зв'язку за критерієм Дарбіна–Уотсона

5. На основі розрахованого значення критерію DW робиться висновок про наявність або відсутність автокореляції залишків:

- якщо $0 < DW < d_L$ – це свідчить про наявність позитивної автокореляції залишків;
- якщо $4 - d_L < DW < 4$ – це свідчить про наявність негативної автокореляції залишків;
- якщо $d_L \le DW \le d_U$ *або* $4 - d_U \le DW \le 4 - d_L$ – неможливо зробити висновок ні про наявність, ні про відсутність автокореляції;
- якщо $d_U < DW < 4 - d_U$, то автокореляції залишків немає.

6.3. Множинна регресія

Часто поведінка параметра Y визначається не одним, а декількома чинниками. Розглянемо випадок, коли шукана залежність лінійна, а кількість визначальних чинників (факторів) – два. Необхідно побудувати рівняння множинної регресії

$$y = a_0 + a_1 x_1 + a_2 x_2 . \tag{6.27}$$

Згідно з принципом Лежандра із умови

$$\sum_{i=1}^{n} \left(y_i - \left(a_0 + a_1 x_{1i} + a_2 x_{2i} \right) \right)^2 \to min \tag{6.28}$$

отримуємо систему лінійних рівнянь для визначення параметрів a_0, a_1, a_2:

$$\begin{cases} na_0 + a_1\sum_{i=1}^{n} x_{1i} + a_2\sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i, \\ a_0\sum_{i=1}^{n} x_{1i} + a_1\sum_{i=1}^{n} x_{1i}^2 + a_2\sum_{i=1}^{n} x_{1i}x_{2i} = \sum_{i=1}^{n} x_{1i}y_i, \\ a_0\sum_{i=1}^{n} x_{2i} + a_1\sum_{i=1}^{n} x_{1i}x_{2i} + a_2\sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i}y_i. \end{cases} \qquad (6.29)$$

Коефіцієнт детермінації R^2, як і у випадку парної регресії, розраховується за формулою: $R^2 = \dfrac{SSR}{SST} = 1 - \dfrac{SSE}{SST}$.

F –статистика розраховується за співвідношенням

$$F = \frac{R^2 / m}{\left(1 - R^2\right)/\left(n - m - 1\right)}, \qquad (6.30)$$

де $(n-m)$ – число ступенів свободи; m – кількість незалежних змінних (у нашому випадку – $m = 2$).

Перевірка статистичної значущості оцінок параметрів багатофакторної лінійної регресії і коефіцієнта множинної кореляції також базується на t–критерії Ст’юдента

$$t_{a_j}^* = a_j / S_{a_j}, \qquad \left(j = \overline{0, m}\right). \qquad (6.31)$$

Теми для роздумів і контрольні запитання

1. Поняття випадкової величини, числові характеристики випадкової величини.

2. Що таке генеральна сукупність, вибірка з генеральної сукупності? Які умови мають виконуватися, щоб вибірка була репрезентативною?
3. За якими формулами розраховують математичне сподівання (чи середнє вибіркове), дисперсію та середнє квадратичне відхилення дискретної випадкової величини?
4. Функціональний та кореляційний зв'язок випадкових величин. Коефіцієнт коваріації.
5. Поняття регресії. Регресійний аналіз. Побудова емпіричних регресійних формул.
6. Визначення коефіцієнтів емпіричної регресійної формули за методом найменших квадратів.
7. Передумови застосування методу найменших квадратів.
8. Лінійна регресія. Знаходження коефіцієнтів лінійної регресійної залежності двох величин.
9. Як провести верифікацію моделі парної лінійної регресії?
10. Що показує коефіцієнт парної кореляції між змінними x і y?
11. Для перевірки статистичної значущості коефіцієнта кореляції і параметрів регресійної моделі використовують критерій…
12. Критерієм чого виступає коефіцієнт детермінації?
13. F –критерій Фішера використовується для перевірки…
14. Поняття довірчого інтервалу. Як побудувати довірчі інтервали для параметрів лінійної регресійної моделі?
15. Точковий та інтервальний прогноз на основі моделі лінійної регресії. Коефіцієнт еластичності.
16. Критерій Дарбіна–Уотсона автокореляції залишків моделі.
17. Множинна лінійна регресія.

7. СТРУКТУРНИЙ АНАЛІЗ СИСТЕМ І ПРОЦЕСІВ

7.1. Дерева рішень

Метод дерев рішень (decision trees) є одним з найбільш популярних методів вирішення задач класифікації та моделювання складних систем. Якщо залежна змінна приймає дискретні значення, то за допомогою методу дерева рішень вирішуються задачі класифікації. Якщо залежна змінна приймає неперервні значення, то дерево рішень вирішує задачу чисельного моделювання та прогнозування. Вперше дерева рішень були запропоновані Ховілендом і Хантом (Hoveland, Hunt) наприкінці 50-х років минулого століття.

У найбільш простому вигляді дерево рішень – це спосіб представлення правил у вигляді ієрархічної структурі. Основа такої структури – відповіді «Так» або «Ні» на низку уточнюючих запитань.

Приклад 1. Побудувати дерево рішень для проблеми: «Чи йти мені грати у футбол?». Щоб прийняти рішення у цьому випадку, слід віднести поточну ситуацію до одного з відомих класів (в даному випадку – «грати» або «не грати»). Для цього потрібно відповісти на низку запитань, які знаходяться у вузлах цього дерева, починаючи з його кореня.

При позитивній відповіді на запитання здійснюється перехід до лівої гілки дерева, при негативній – до правої гілки дерева (рис. 7.1). Далі йде наступне запитання і так, поки не буде досягнуто кінцевий вузол дерева, що є вузлом рішення. Можливі ланцюжки прийняття рішень будуть такими.

1. Корінь дерева: «Сонячно?».
   Відповідь – «Так».
   Внутрішній вузол дерева: «Температура повітря висока?». Відповідь – «Так».
   Рішення – «Не грати».
2. Корінь дерева: «Сонячно?».
   Відповідь – «Ні».
   Внутрішній вузол дерева: «Чи йде дощ?». Відповідь – «Ні».
   Рішення – «Грати».

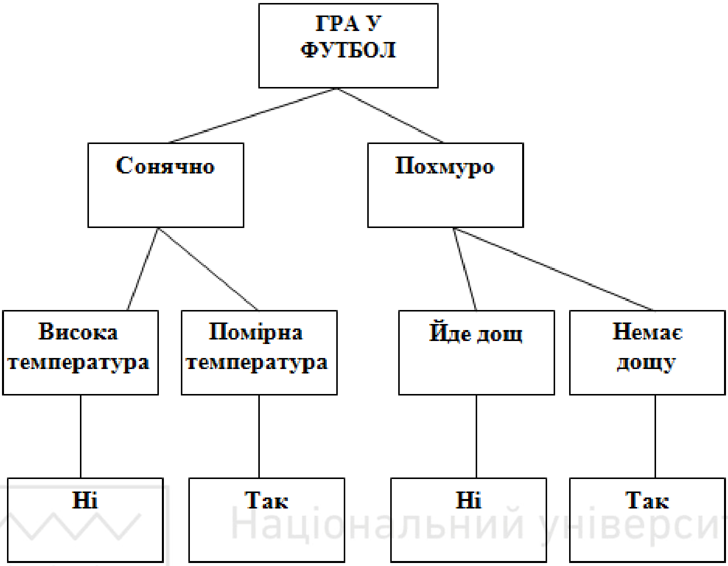


Рис. 7.1. Структурний аналіз проблеми «Гра у футбол»

У розглянутому прикладі вирішується задача бінарної класифікації. У вузлах бінарних дерев розгалуження може вестися тільки в двох напрямках (відповіді «так» і «ні»). У загальному випадку відповідей і, відповідно, гілок дерева, що виходять з його внутрішнього вузла, може бути більше двох.

Приклад 2. Побудувати дерево рішень для ситуації «Чи йти мені на вечірку?»

Вирішуючі питання:

Чи вмію я танцювати?

Чи я не втомлений?

Чи готовий я до завтрашніх занять?

Чи йде на вечірку мій товариш (подружка)?

Приклад 3. **Аналіз кредитоспроможності клієнта.** База даних, на основі якої повинно здійснюватися прийняття рішення, містить такі дані про клієнтів банку: вік, наявність нерухомості, освіта, середньомісячний дохід, чи вчасно повертав клієнт раніше отримані кредити. Завдання полягає в тому, щоб на підставі перелічених вище даних визначити, чи варто видавати кредит цьому клієнтові.

Задача вирішується в два етапи: побудова класифікаційної моделі і її використання. На етапі побудови моделі будується дерево класифікації або створюється набір якихось правил. На етапі використання моделі побудоване дерево, яке є експертною системою для перевірки конкретного клієнта, використовується для відповіді на поставлене запитання «Чи видавати кредит?».

Класифікаційні правила будуються у вигляді конструкцій «якщо – то».

Приклади правил:

Якщо «Вік < 35» то слід видати кредит.

Якщо «Вік > 35» і «Дохід > 7000», то слід видати кредит.

Якщо «Вік > 60» і «Є нерухомість», то слід видати кредит.

Якщо «Вік > 60» і «Немає нерухомості» кредит видавати не слід.

Якщо «Попередній кредит повернутий невчасно» – кредит видавати не слід.

Переваги дерев рішень

- Класифікаційна модель, представлена у вигляді дерева рішень, розбиває рішення на взаємовиключні варіанти і спрощує розуміння розв'язуваної задачі.
- На побудову класифікаційних моделей за допомогою алгоритмів конструювання дерев рішень потрібно значно менше часу, ніж на навчання нейронних мереж.
- Дерева рішень дають можливість будувати правила з бази даних на природній мові.
- Дерева рішень працюють як з числовими, так і з номінальними типами даних.

7.2. Структурний аналіз дерева подій

Складна система має багато варіантів поведінки. Якщо система добре вивчена, кожен варіант можна охарактеризувати певною ймовірністю реалізації та очікуваним фінансовим результатом. Детальний аналіз усіх можливих варіантів поведінки дозволяє приймати оптимальні управлінські рішення.

При аналізі бізнес-проектів розглядають так зване «дерево подій», утворене в результаті аналізу альтернативних варіантів розвитку ситуації. Розглянемо деякі приклади побудови дерева подій, які ілюструють методику оцінювання різних варіантів поведінки системи.

Приклад 4. **Аналіз дохідності бізнес проекту.** Розглянемо схему деякого бізнес-процесу (рис. 7.2). Процес може розвиватися за двома варіантами: позитивний варіант – збільшення продаж (світлий прямокутник, прибуток 200 тис. грн.) та негативний варіант – зменшення продаж (темний прямокутник, прибуток 100 тис. грн.). Ймовірність кожного варіанту, визначена експертами, вказана всередині прямокутника. Очікуваний прибуток дорівнює добутку ймовірності відповідного варіанту на його прибутковість.

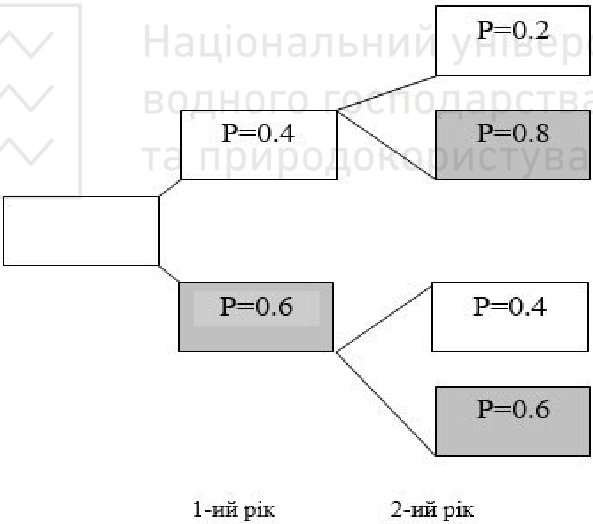


Рис. 7.2. Структурний аналіз дохідності бізнес-проекту

Розрахуємо очікуваний прибуток за перший рік реалізації проекту

$$Z = p_1 \cdot z_1 + p_2 \cdot z_2 = 200 \cdot 0{,}4 + 100 \cdot 0{,}6$$

$$Z = 80 + 60 = 140 \text{ (тис. грн.)}$$

На другому році реалізації проекту можливі 4 варіанти. Їх ймовірність визначається за правилом ймовірності складних подій.

Складна подія – це подія B, яка відбулася за умови, що перед нею відбулася інша подія A. Відомо, що події A і B є незалежними. Позначимо $P(A)$ – ймовірність реалізації події A, $P(B)$ – ймовірність реалізації події B, $P(B|A)$ – умовна ймовірність реалізації події *B,* у випадку, що перед нею мала місце подія *A*. Тоді має місце співвідношення

$$P(B|A) = P(A) \cdot P(B). \tag{7.1}$$

Формулу (7.1) називають формулою ймовірності складної події.

Таким чином, ймовірність першого варіанту подій (рис. 7.2, 2 рік) дорівнює

$$P = 0{,}4 \cdot 0{,}2 = 0{,}08.$$

Ймовірність другого варіанту подій дорівнює:

$$P = 0{,}4 \cdot 0{,}8 = 0{,}32.$$

Ймовірність третього варіанту подій дорівнює:

$$P = 0{,}6 \cdot 0{,}4 = 0{,}24.$$

Ймовірність четвертого варіанту подій дорівнює

$$P = 0{,}6 \cdot 0{,}6 = 0{,}36.$$

Сума ймовірностей всіх подій дорівнює 1, тобто розглянуті нами 4 варіанти розвитку події утворюють повну групу подій.

Повна група подій – це такий перелік можливих подій, одна з яких має обов'язково відбутися. Нехай система може перебувати в n різних несумісних станах $S_1, S_2, \ldots, S_n$. Ймовірність кожного із станів відома і становить $P_1, P_2, \ldots, P_n$. Якщо виконується рівність

$$P_1 + P_2 + \ldots + P_n = 1, \qquad (7.2)$$

то група станів $S_1, S_2, \ldots, S_n$ називається повною.

Розрахуємо очікуваний прибуток за другий рік реалізації проекту

$$Z = p_1 \cdot z_1 + p_2 \cdot z_2 + p_3 \cdot z_3 + p_4 \cdot z_4$$

$$Z = 400 \cdot 0{,}08 + 300 \cdot 0{,}32 + 300 \cdot 0{,}24 + 200 \cdot 0{,}36$$

$$Z = 32 + 96 + 72 + 72 = 272 \text{ (тис. грн.)}$$

Приклад 5. **Аналіз варіантів здобуття професії.**

Випускник школи планує поступати в університет або у коледж. Відповідні ймовірності становлять 0,7 (університет) та 0,3 (коледж). Після закінчення коледжу випускник планує працювати референтом (зарплата 1500 грн).

У разі вступу до університету випускник планує навчатися за напрямом «Економічна кібернетика» (ймовірність 0,6) або ж за напрямом «Документознавство» (ймовірність 0,4). Після закінчення навчання за напрямом «Документознавство» випускник планує працювати у відділі кадрів (зарплата 2500 грн).

У випадку закінчення навчання за напрямом «Економічна кібернетика» випускник планує влаштуватися на роботу у страхову компанію (зарплата 4000 грн, ймовірність 0,45) або ж у банк (зарплата 5000 грн, ймовірність 0,4) чи працювати програмістом (ймовірність 0,15).

У разі роботи програмістом розглядаються два варіанти: робота в Україні (зарплата 10000 грн, ймовірність 0,75) та робота за кордоном (зарплата 35000 грн, ймовірність 0,25).

Для економічної оцінки перспектив розглянутих варіантів слід враховувати як економічну вигоду так і ймовірність відповідного варіанту. Якщо подія є результатом ланцюжка попередніх подій, то це – складна подія і її ймовірність оцінюють як добуток ймовірностей всіх попередніх подій. На кожному рівні еволюції процесу слід перевіряти формулу повної групи подій (сума ймовірностей дорівнює одиниці).

Структурний аналіз проблеми зображено на рис. 7.3. Структурний аналіз складається з послідовності наступних етапів.

Етап 1. Економічна оцінка перспективності варіанту «вступ до коледжу» визначається як добуток ймовірності та очікуваного доходу

$$Z = p_1 \cdot z_1 = 1500 \cdot 0{,}3 = 450 \text{ (грн).}$$

Етап 2. З використанням формули складної ймовірності визначаємо економічну оцінку перспективності варіанту «університет» → «Документознавство»

$$Z = p_2 \cdot z_2 = 2500 \cdot 0{,}7 \cdot 0{,}4 = 700 \text{ (грн).}$$

Етап 3. Визначаємо економічну оцінку перспективності варіанту «університет» → «Економічна кібернетика» → «Страхова компанія»

$$Z = 4000 \cdot 0{,}7 \cdot 0{,}6 \cdot 0{,}45 = 756 \text{ (грн).}$$

Визначаємо економічну оцінку перспективності варіанту «університет» → «Економічна кібернетика» → «Банк»

$$Z = 5000 \cdot 0{,}7 \cdot 0{,}6 \cdot 0{,}4 = 840 \text{ (грн).}$$

Етап 4. Визначаємо економічну оцінку перспективності варіанту «університет» → «Економічна кібернетика» → «Програміст» → «Україна»

$$Z = 10000 \cdot 0{,}7 \cdot 0{,}6 \cdot 0{,}15 \cdot 0{,}7 = 441 \text{ (грн).}$$

Рис. 7.3. Структурний аналіз проблеми «Здобуття професії»

Визначаємо економічну оцінку перспективності варіанту «університет» $\rightarrow$ «Економічна кібернетика» $\rightarrow$ «Програміст» $\rightarrow$ «за кордон»

$$Z = 35000 \cdot 0{,}7 \cdot 0{,}6 \cdot 0{,}15 \cdot 0{,}3 = 661 \text{ (грн).}$$

Висновок. З усіх переглянутих варіантів найбільш економічно перспективним є варіант «університет» $\rightarrow$ «Економічна кібернетика» $\rightarrow$ «Банк».

7.3. Структурний аналіз дерева відмов

Аналіз дерева відмов вважається одним з найбільш корисних аналітичних інструментів у процесі оцінювання безпеки експлуатації складних систем. Цей метод використовує дедуктивний логічний метод (рух від загального до часткового). Зазвичай, небажані події відбуваються під впливом різних чинників. Дерево відмов дозволяє проаналізувати дію цих чинників на різних етапах роботи системи. При аналізі дерева відмов небажану подію (аварія, банкрутство) вважають кінцевою точкою схеми. Розглядаючи окремі події, які передували кінцевій події, будують дерево відмов. Прикладом дерева відмов є аналіз пасажирських перевезень з урахуванням імовірності поломок автобуса.

Приклад 6. **Системний аналіз пасажирських перевезень**

Методику системного аналізу розглянемо на прикладі аналізу маршруту системи міжміських пасажирських перевезень.

Основні характеристики маршруту

Ціна нового автобуса – 300 тис. грн. Ціна автобуса, який був у використанні (б/в) – 150 тис. грн.

Орієнтовний річний дохід – 700 тис. грн.

Вартість ремонту двигуна – 40 тис. грн. Встановлення нового двигуна – 100 тис. грн. Після 2-ої поломки двигуна (для б/в двигуна – після 1-ої поломки) – заміна двигуна.

Річні експлуатаційні витрати: новий автобус – 300 тис. грн, б/в автобус – 450 тис. грн.

Стартовий капітал підприємця – 300 тис. грн.

Завдання. Скласти бізнес-план роботи на 3 роки, підрахувавши очікуваний прибуток для кожного року.

Таблиця 7.1

Ймовірність поломки двигуна

	1-ий рік	2-ий рік	3-ий рік
Новий	0	0,1	0,3
Після ремонту (або б/в)	0,1	0,3	0,6

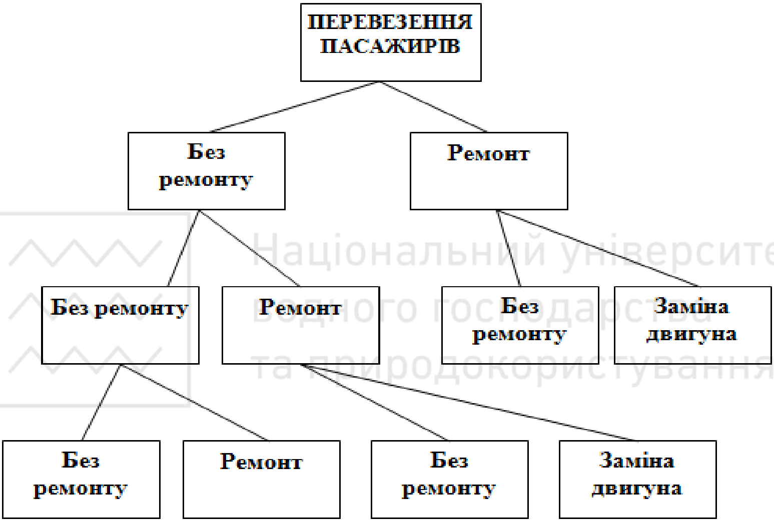


Рис. 7.4. Структурний аналіз задачі «Поломка двигуна»

Система може перебувати у двох станах: S_1 – двигун працює; S_2 – двигун не працює. Ймовірність першого стану позначимо p, ймовірність другого q. Оскільки події S_1 і S_2 утворюють повну групу подій, маємо рівність $p+q=1$.

Наприклад, ймовірність безаварійної роботи нового двигуна на другому році експлуатації $P(A)=0{,}9$, ймовірність безаварійної роботи нового двигуна на третьому році експлуатації $P(S_1)=0{,}7$. Якщо розглядати ці події як незалежні, то ймовірність безаварійної роботи двигуна на третьому році експлуатації за умови безаварійної роботи на другому році становить

$$P(S_i \mid A) = P(A) \cdot P(S_i) = 0{,}9 \cdot 0{,}7 = 0{,}63\,.$$

Ймовірність варіанту, при якому на другому році експлуатації була поломка двигуна, а третій рік пройшов без аварій становить

$$P(S_i \mid \overline{A}) = P(\overline{A}) \cdot P(S_i) = 0{,}1 \cdot (1 - 0{,}1) = 0{,}09\,.$$

Звертаємо увагу на те, що після ремонту двигуна ймовірність поломки слід розрахувати за даними другого рядка табл. 7.1 (двигун після ремонту). Ці ж дані слід використовувати у випадку автобуса, який був у використанні.

Варіант 1. Новий автобус

При аналізі експлуатації нового автобуса протягом 3-х років необхідно розглянути різні варіанти, які можуть скластися у залежності від надійності роботи його двигуна. При цьому слід використовувати поняття складної події та її ймовірності. Слід також пам'ятати, що на першому році експлуатації поломка нового двигуна практично виключена. У залежності від надійності роботи двигуна на 2-му та 3-му році експлуатації можливі 4 варіанти експлуатації:

1. $БР_2 + БР_3$. Двигун служить без ремонту протягом трьох років експлуатації. Ймовірність такого варіанту становить $P_1 = 0{,}9 \cdot 0{,}7 = 0{,}63$. Затрати при цьому варіанті дорівнюють $300 \cdot 3 = 900$ тис. грн.
2. $БР_2 + Р_3$. Другий рік без ремонту, третій рік – ремонт. Ймовірність такого варіанту становить $P_2 = 0{,}9 \cdot 0{,}3 = 0{,}27$. Затрати при даному варіанті становлять $300 \cdot 3 + 40 = 940$ тис. грн.
3. $Р_2 + БР_3$. Ремонт на другому році і безремонтний третій рік. Ймовірність такого варіанту становить

$P_3 = 0{,}1 \cdot 0{,}9 = 0{,}09$. Затрати при даному варіанті становлять $300 \cdot 3 + 40 = 940$ тис. грн.

4. $P_2 + P_3$. Ремонт на другому році та заміна двигуна на третьому році. Ймовірність такого варіанту становить $P_4 = 0{,}1 \cdot 0{,}1 = 0{,}01$. Затрати при даному варіанті становлять $300 \cdot 3 + 40 + 100 = 1040$ тис. грн. Звертаємо увагу, що оскільки повторний ремонт двигуна не рекомендується, на третьому році необхідно встановити новий двигун.

Середньоймовірнісні затрати трьохрічної експлуатації становлять

$$Z = P_1 \cdot Z_1 + P_2 \cdot Z_2 + P_3 \cdot Z_3 + P_4 \cdot Z_4$$

$$Z = 0{,}63 \cdot 900 + 0{,}27 \cdot 940 + 0{,}09 \cdot 940 + 0{,}01 \cdot 1040$$

$$Z = 567{,}0 + 253{,}8 + 84{,}6 + 10{,}4 = 915{,}8 \text{ (тис. грн.)}$$

Очікуваний дохід становить

$$700 \cdot 3 = 2100 \text{ тис. грн.}$$

Очікуваний прибуток становить

$$2100 - 915{,}8 = 1184{,}2 \text{ тис. грн.}$$

Отже, як показали розрахунки, очікуваний прибуток від експлуатації нового автобуса за 3 роки становить 1184,2 тис. грн.

Варіант 2. Автобус, який був у використанні

Експлуатація двигуна автобуса, який вже був у використанні, прирівнюється до експлуатації двигуна після першого ремонту. Тому, вже після першої поломки такого двигуна необхідно встановлювати новий двигун. Після заміни двигуна, двигун вважається новим і ймовірність його поломки розраховується з використанням першого рядка табл. 7.1.

Ймовірність безвідмовної роботи такого двигуна на першому році експлуатації становить 1, на другому році – 0,9, на третьому році – 0,7. За початковий капітал 300 тис. грн. можна купити 2 автобуси БВ. При аналізі експлуатації автобуса БВ протягом 3-х років необхідно розглянути різні варіанти, які можуть скластися у залежності від надійності роботи двигуна (табл. 7.1). Всього існує 8 варіантів експлуатації автобуса:

1. $БР_1 + БР_2 + БР_3$. Двигун служить без ремонту протягом трьох років експлуатації. Ймовірність такого варіанту становить $P_1 = 0{,}9 \cdot 0{,}7 \cdot 0{,}4 = 0{,}252$. Затрати при цьому дорівнюють $Z_1 = 450 \cdot 3 = 1350$ тис. грн.
2. $БР_1 + БР_2 + P_3$. Перший і другий рік без ремонту, третій рік – ремонт (заміна двигуна). Ймовірність такого варіанту становить $P_2 = 0{,}9 \cdot 0{,}7 \cdot 0{,}6 = 0{,}378$. Затрати при цьому дорівнюють $Z_2 = 450 \cdot 3 + 100 = 1450$ тис. грн.
3. $БР_1 + P_2 + БР_3$. Ремонт на 2-му році і безремонтні 1-ий і 3-й роки. Ймовірність такого варіанту становить $P_3 = 0{,}9 \cdot 0{,}3 \cdot 1 = 0{,}270$. Затрати при цьому варіанті дорівнюють $Z_3 = 450 \cdot 3 + 100 = 1450$ тис. грн. Звертаємо увагу на те, що ймовірність поломки двигуна на третьому році дорівнює нулю, оскільки двигун перед цим був замінений.
4. $БР_1 + P_2 + P_3$. Ремонт на 2-му році і поломка на 3-му році. Ймовірність такого варіанту дорівнює нулю (ймовірність поломки двигуна на третьому році дорівнює нулю, оскільки двигун перед цим був замінений) $P_4 = 0{,}9 \cdot 0{,}3 \cdot 0 = 0$.
5. $P_1 + БР_2 + БР_3$. Заміна двигуна на першому році і безремонтні 2-ий і 3-й роки. Ймовірність такого варіанту становить $P_5 = 0{,}1 \cdot 1 \cdot 0{,}9 = 0{,}090$. Затрати при цьому дорівнюють $Z_5 = 450 \cdot 3 + 100 = 1450$ тис. грн.

6. $P_1 + БP_2 + P_3$. Заміна двигуна на першому році і ремонт на 3-му році. Ймовірність такого варіанту становить $P_6 = 0{,}1 \cdot 1 \cdot 0{,}1 = 0{,}010$. Затрати при цьому варіанті дорівнюють $Z_6 = 450 \cdot 3 + 40 + 100 = 1490$ тис. грн.
7. $P_1 + P_2 + БP_3$. Заміна двигуна на першому році і ремонт двигуна на 2-му році. Ймовірність такого варіанту становить $P_7 = 0$.
8. $P_1 + P_2 + P_3$. Ймовірність даного варіанту також нульова. $P_8 = 0$.

Середньоймовірнісні витрати трьохрічної експлуатації становлять

$$Z = P_1 \cdot Z_1 + P_2 \cdot Z_2 + P_3 \cdot Z_3 + P_4 \cdot Z_4 + P_5 \cdot Z_5 + \\ + P_6 \cdot Z_6 + P_7 \cdot Z_7 + P_8 \cdot Z_8$$

$$Z = 0{,}252 \cdot 1350 + 0{,}378 \cdot 1450 + 0{,}270 \cdot 1450 + 0 \cdot 1490 + \\ + 0{,}090 \cdot 1450 + 0{,}010 \cdot 1490 + 0 \cdot 1550 + 0 \cdot 1650$$

$$Z = 340{,}2 + 548{,}1 + 391{,}5 + 130{,}5 + 14{,}9 = 1425{,}2 \text{ (тис. грн.)}$$

Очікуваний дохід становить

$$700 \cdot 3 \cdot 2 = 4200 \text{ тис. грн.}$$

Очікуваний прибуток становить

$$(2100 - 1425{,}2) \cdot 2 = 1349{,}6 \text{ тис. грн.}$$

Висновок. Другий варіант бізнес-плану є більш вигідним.

7.4. Задача про видачу кредиту

Бізнесмен просить банк про кредит $15000. Банк приймає рішення з двох альтернатив:

1. Видати кредит під 15% річних. Статистика свідчить, що 4% таких кредитів не повертаються (останній рядок табл. 7.2).
2. Вкласти ці гроші ($15000) у інший бізнес-проект з гарантованим поверненням під 9% річних.
3. Перед видачею кредиту можна перевірити кредитоспроможність клієнта в аудиторській фірмі. Ця послуга коштує $80.

Статистику раніше виданих кредитів наведено в табл. 7.2.

Таблиця 7.2

Статистика виданих кредитів

Рекомендація після перевірки	Фактичний результат		
	Кредит повернуто	Кредит не повернуто	Всього
Видати кредит	735	15	750
Не видавати кредит	225	25	250
Всього	**960**	**40**	**1000**

Побудова дерева рішень. Побудуємо дерево рішень (рис. 7.5). Пунктирні лінії з'єднують квадрати можливих рішень. Суцільні лінії з'єднують круги можливих результатів. Квадратні вузли позначають моменти прийняття рішень. Круглі вузли позначають отримані результати. Для кожної вітки дерева необхідно прорахувати її ймовірність та очікуваний фінансовий результат.

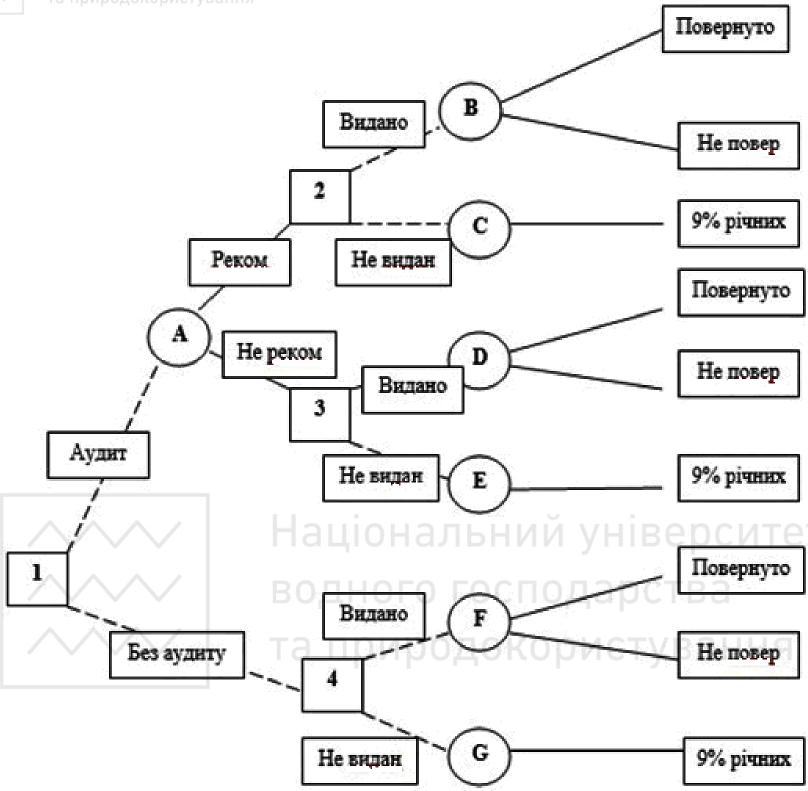


Рис. 7.5. Структурний аналіз проблеми про видачу кредиту

Розрахунок ймовірностей. Спочатку розрахуємо ймовірності всіх варіантів, використовуючи статистичні дані.

1. Фірма рекомендувала, кредит повернуто

$$P = \frac{735}{750} = 0{,}98\,.$$

2. Фірма рекомендувала, кредит не повернуто

$$P = \frac{15}{750} = 0{,}02\,.$$

3. Фірма не рекомендувала, кредит повернуто

$$P = \frac{225}{250} = 0{,}90\,.$$

4. Фірма не рекомендувала, кредит не повернуто

$$P = \frac{25}{250} = 0{,}10\,.$$

Аналіз дерева рішень

1. Спочатку розглянемо варіанти прийняття рішення, які опираються на поради аудитора. Розглядаємо вітки В і С, які є наслідками рекомендації «видати кредит». Приймаємо рішення В (видати кредит).

Очікуваний дохід при події В: $(15000 + 2250)\cdot 0{,}98 - 15000\cdot 0{,}02 = 16605$.

Чистий дохід: $16605 - 15000 = 1605$.

Розглянемо рішення С (не видавати кредит). Хоча це рішення і суперечить пораді аудитора, але кінцеве рішення приймає банк. Оскільки кредит не видається, ця сума коштів ($15000) інвестується у безризиковий проект з дохідністю 9% річних.

Очікуваний дохід при події С: $(15000 + 1350)\cdot 1{,}00 = 16350$.

Чистий дохід: $16350 - 15000 = 1350$.

Отже, перебуваючи у квадраті 2 слід вибрати варіант В, оскільки він має більший очікуваний дохід. Приймаємо рішення: варіант С відхилити (перекреслити вітку). Квадрату 2 присвоїти очікуваний дохід 1605.

2. Тепер розглянемо вітки D і E, які є наслідками рекомендації аудитора 3 (не рекомендувати видати кредит).

Приймаємо рішення D (видати кредит). Очікуваний дохід при події D становить: $17250 \cdot 0{,}90 - 15000 \cdot 0{,}10 = 14025$. Чистий дохід: $14025 - 15000 = -975$.

Розглянемо рішення E (не видавати кредит). Очікуваний дохід при події E: 16350. Чистий дохід: $16350 - 15000 = 1350$.

Таким чином, перебуваючи у квадраті 3 слід вибрати варіант E, оскільки він має більший очікуваний дохід. Приймаємо рішення: варіант D відхилити (перекреслити вітку). Квадрату 3 присвоїти очікуваний дохід 1350.

3. Тепер розглянемо вітки F і G, які є наслідком рішення 4 (рішення приймається без аудиту).

Рішення F (видати кредит). Очікуваний дохід при події F: $17250 \cdot 0{,}96 - 15000 \cdot 0{,}04 = 15960$.

Чистий дохід: $15960 - 15000 = 960$.

Рішення G (не видавати кредит). Очікуваний дохід при події G: 16350. Чистий дохід: $16350 - 15000 = 1350$.

Отже, перебуваючи у квадраті 4 слід вибрати варіант G, оскільки він має більший очікуваний дохід. Приймаємо рішення: варіант F відхилити (перекреслити вітку). Квадрату 4 присвоїти очікуваний дохід 1350.

4. Тепер порівняємо вузли A і 4. Розрахуємо математичне сподівання для вузла A: $1605 \cdot 0{,}75 + 1350 \cdot 0{,}25 = 1541{,}25$.
Чистий дохід: $1541{,}25 - 80 = 1461{,}25$.

Отже, перебуваючи у квадраті 1 слід вибрати варіант A, оскільки він має більший очікуваний дохід, ніж варіант 4 (1461,25 > 1350). Отже, приймаємо рішення: варіант 4 відхилити (перекреслити вітку). Квадрату 1 присвоїти очікуваний дохід 1461,25.

Висновок. Оптимальне рішення відповідає гілці дерева рішень: 1 – A. Якщо аудитор рекомендує видати кредит, то слід видати кредит, у протилежному випадку – кредит не видавати, а інвестувати гроші у безризиковий проект.

Теми для роздумів і контрольні запитання

1. Поняття дерева рішень. Бінарні (класифікаційні) дерева рішень. Наведіть приклади.
2. У чому полягають переваги і недоліки структуризації проблеми у вигляді бінарного дерева рішень?
3. Чи будь-яку проблему можна представити у вигляді бінарного дерева рішень?
4. Що таке «дерево подій»? Наведіть приклади.
5. Що таке випадкова подія, складна подія? Як розраховується ймовірність складної випадкової події?
6. Поняття повної групи подій.
7. Для чого використовується «дерево відмов»? Наведіть приклади.
8. Загальний вигляд дерева рішень бізнес-процесу. Основні структурні елементи і процес побудови.
9. Використання дерева рішень бізнес-процесу для розрахунку очікуваних фінансових результатів.

8. СИСТЕМНИЙ АНАЛІЗ ТРАНСПОРТНИХ МЕРЕЖ

8.1. Основні поняття теорії графів

Теорія графів – одна з важливих частин математичного апарату інформатики та кібернетики. У термінах теорії графів можна сформулювати багато задач, пов’язаних із дискретними об’єктами. Велику роль відіграють алгоритми на графах. Теорія графів являє собою дуже зручну мову для опису програмних та багатьох інших моделей. Струнка система спеціальних термінів та означень теорії графів дозволяє просто і доступно описувати складні та тонкі речі. Особливо важливим є наявність наочної графічної інтерпретації поняття графу. Зображення дозволяють відразу побачити суть питання на інтуїтивному рівні.

Зокрема, використовуючи математичний апарат теорії графів можна ефективно оптимізувати управління багатьма системами. Як *приклад,* можна говорити про управління водопровідними, електричними, транспортними та іншими мережами, зокрема, міського господарства чи ін.

З поняттям графу зазвичай пов’язують його графічне представлення, при якому він зображується як множина точок, деякі з яких з’єднані лініями. Але граф відрізняється від геометричних конфігурацій (скажімо, фігур, які також складаються з точок та ліній) тим, що в графі несуттєві відстані між точками, форма з’єднувальних ліній та кути між ними. Важливо лише те, чи з’єднана дана пара точок лінією, чи ні. Тому граф іноді називають топологічним об’єктом, тобто об’єктом, властивості якого не змінюються при розтягуванні, стисненні та викривленні. З цієї ж причини (важливим є тільки наявність або відсутність з’єднань) граф – об’єкт дискретний і може бути заданий двома дискретними множинами: множиною точок (вершин), та множиною ліній (ребер), які з’єднують деякі вершини.

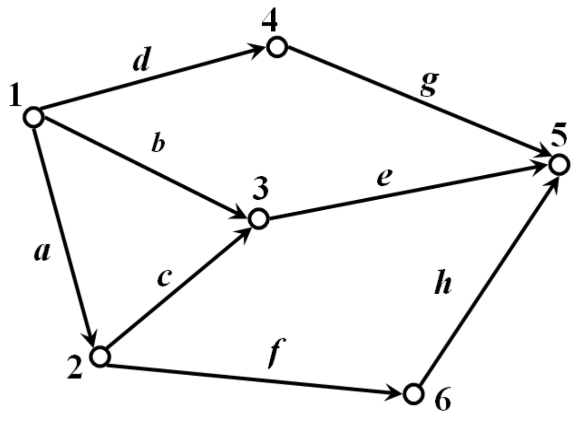
Графом $G(V,E)$ називається система, яка складається з двох множин: V – множини ***вершин*** (*vertices*) і E – множини ***ребер*** (*edges*), що з’єднують вершини. Кількість вершин графу позначають n і називають ***порядком*** графу.

Вершина і ребро називають ***інцидентними***, якщо ця вершина є одним з кінців ребра. Дві вершини, які з'єднані ребром, називають ***суміжними***. Ребра, які мають спільну вершину, теж називають ***суміжними***.

Скінченна послідовність суміжних ребер називається ***маршрутом***. Маршрут, у якому всі ребра різні називається ***ланцюгом***. Ланцюг, у якого всі вершини різні, називається ***простим ланцюгом***. Граф G називається *зв'язним*, якщо будь-які дві його вершини можна сполучити хоча б одним ланцюгом. Ланцюг, у якого початкова і кінцева вершини збігаються, називається ***циклом***.

Зв'язний граф, який не має циклів, називається ***деревом***. Для кожного з графів можна побудувати дерево, яке містить всі його вершини. Таке дерево називають *покриваючим*, або *покриттям графа*. Якщо кожне ребро графа має напрямок, граф називають *орієнтованим* (***орграфом***). Напрямлене ребро орграфа називають ***дугою***. ***Мережею*** називають орграф, кожна дуга якого має певне числове значення (вагу).

Граф можна задавати у вигляді рисунка (рис. 8.1, а) чи переліком вершин і ребер (рис. 8.1, б). Проте найзручніший спосіб задання графа для комп'ютерної обробки – представлення графа за допомогою матриць інцидентності чи суміжності.



а)

Ребро	Вершини
a	1, 2
b	1, 3
c	2, 3
d	1, 4
e	3, 5
f	2, 6
g	4, 5
h	5, 6

б)

Рис. 8.1. Приклад орієнтованого графа

Наприклад, граф, поданий на рис. 8.1, може бути заданий за допомогою матриці інцидентності (8.1). Кожен рядок цієї матриці відповідає вершині, а стовпець – дузі. У кожному стовпці такої матриці є один елемент «1» (вхід), один елемент «–1» (вихід), а решта елементів дорівнюють нулю.

$$\begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}. \tag{8.1}$$

Використовуючи матриці інцидентності або матриці суміжності можна отримати і інші матриці, які відповідають графу та містять багато важливої інформації про граф у зручній формі, зокрема, *матриці відстаней* і *матриці досяжності*.

Кількість ребер, які інцидентні вершині *v*, називається ***степенем*** (або ***валентністю***) вершини *v* і позначається $d(v)$. Степінь вершин легко порахувати за матрицею інцидентності. Справді, в *i*-му рядку матриці інцидентності, який відповідає вершині v_i, одиниці знаходяться на перетині зі стовпцями, яким відповідають інцидентні цій вершині ребра, а інші елементи рядка дорівнюють 0. Сума одиниць рядка і буде степенем відповідної вершини.

Для орграфу кількість дуг, які виходять з вершини *v*, називається ***напівстепенем виходу***, а вхідних – ***напівстепенем входу***. Позначаються вони, відповідно, $d^-(v)$ та $d^+(v)$.

Теорема Ейлера. Сума степенів вершин графу дорівнює подвоєній кількості його ребер (*m*): $\sum_{i=1}^{n} d(v_i) = 2m$, а сума напівстепенів вузлів орграфу дорівнює подвійній кількості дуг (m):

$$\sum_{i=1}^{n} d^{-}(v_i) + \sum_{i=1}^{n} d^{+}(v_i) = 2m .$$

У нашому *прикладі* (рис. 8.1): $d^{-}(1) = 3$, $d^{+}(1) = 0$, $d^{-}(2) = 2$, $d^{+}(2) = 1$, $d^{-}(6) = 1$, $d^{+}(6) = 1$.

8.2 Відстань між вершинами графу. Хвильовий алгоритм

Для зважених графів (мереж) ***довжиною*** маршруту (шляху) є сума ваг ребер (дуг) цього маршруту (шляху). У незважених графах вага кожного ребра (дуги) вважається рівною 1. Тому, за ***довжину*** маршруту (шляху) у таких графах приймають кількість ребер (дуг) цього маршруту (шляху).

Розглянемо деякі властивості мінімальних маршрутів (шляхів) для незважених графів.

Назвемо ***образом вершини*** v в орієнтованому графі G множину кінців дуг, початком яких є вершина v (позначається $D(v)$), а множину початків дуг, кінцем яких є вершина v, назвемо прообразом вершини v (позначається $D^{-1}(v)$). Зрозуміло, що $D(v) \cap D^{-1}(v) = \Gamma(v)$, де $\Gamma(v)$ – ***множина суміжності*** вершини v.

Нехай $G = (V, E)$ – орієнтований граф з n вершинами ($n \geq 2$), а v, u – задані вершини з V, де $v \neq u$. Опишемо алгоритм пошуку відстані та відповідного їй мінімального шляху з v до u в орієнтованому графі G. Цей алгоритм має назву ***хвильового***.

Алгоритм (хвильовий)

1. Позначимо вершину v індексом 0, а вершини, що належать образу вершини v, – індексом 1. Множину вершин з індексом k позначимо $F_k(v)$. Вважаємо, що k=1.

2. Якщо $F_k(v)=\varnothing$ або виконується $k=n-1$ і $u \notin F_k(v)$, то вершина u є незв'язаною з v і робота алгоритму завершується. В іншому випадку – перейти до пункту 3.

3. Якщо $u \notin F_k(v)$, то переходимо до пункту 4. В іншому випадку – існує шлях з v до u завдовжки k, причому цей шлях є мінімальним. Послідовність вершин $v, u_1, u_2, \ldots, u_{k-1}, u,$ де

$$u_{k-1} \in F_{k-1}(v) \cap D^{-1}(u),$$
$$u_{k-2} \in F_{k-2}(v) \cap D^{-1}(u_{k-1}),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$u_1 \in F_1(v) \cap D^{-1}(u_2),$$

і є шуканим мінімальним шляхом з v у u. На цьому робота алгоритму завершується.

4. Позначимо індексом k+1 всі непозначені вершини, які належать образу множини вершин з індексом k. Множину вершин з індексом k+1 позначимо $F_{k+1}(v)$. Збільшуємо індекс k на 1 і переходимо до пункту 2.

Назва алгоритму – хвильовий – пов'язана з тим, що визначення індексів k вершин графу G відбувається, як розповсюдження з початкової вершини v певної хвилі, яка спрямовується за напрямком дуг. Коли хвиля дійде до кінцевої вершини u, це буде означати завершення алгоритму. Значення індексу, «принесеного хвилею», у вершині u буде відповідати довжині знайденого маршруту. А для того, щоб визначити цей маршрут (послідовність вершин), потрібно з кінцевої вершини u повертатися у зворотному до розповсюдження хвилі напрямку і відзначати послідовно одну довільну вершину із значеннями індексу $k-1$, $k-2$, …, 1, 0. Зрозуміло, що вершина з індексом 0 – це початкова вершина v.

Вершини u_1, $u_2, \ldots,$ u_{k-1} можуть бути визначені неоднозначно. Ця неоднозначність відповідає випадкам, коли в орграфі G існує кілька різних мінімальних шляхів з v до u.

Наприклад, визначимо мінімальний шлях з v_1 до v_6 в орієнтованому графі G, заданому матрицею суміжності (рис. 8.2).

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	0	0	1	1	0
v_2	1	0	0	0	0	1
v_3	0	1	0	0	0	1
v_4	0	1	0	0	1	0
v_5	1	0	1	0	0	0
v_6	0	0	1	0	1	0

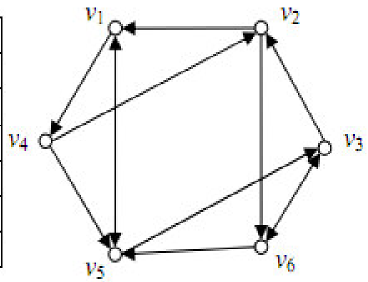


Рис. 8.2. Хвильовий алгоритм

За хвильовим алгоритмом послідовно знаходимо:

$F_1(v_1) = \{v_4, v_5\}$;
$F_2(v_1) = D(F_1(v_1)) \setminus \{v_1, v_4, v_5\} = \{v_2, v_3\}$;
$F_3(v_1) = D(F_2(v_1)) \setminus \{v_1, v_2, v_3, v_4, v_5\} = \{v_6\}$.

Національний університет
водного господарства
та природокористування

Оскільки, $v_6 \in F_3(v_1)$, то за пунктом 3 існує шлях з v_1 до v_6 завдовжки 3, і цей шлях є мінімальним.

Знайдемо тепер цей мінімальний шлях із з v_1 до v_6. Визначимо множину

$$F_2(v_1) \cap D^{-1}(v_6) = \{v_2, v_3\} \cap \{v_2, v_3\} = \{v_2, v_3\}.$$

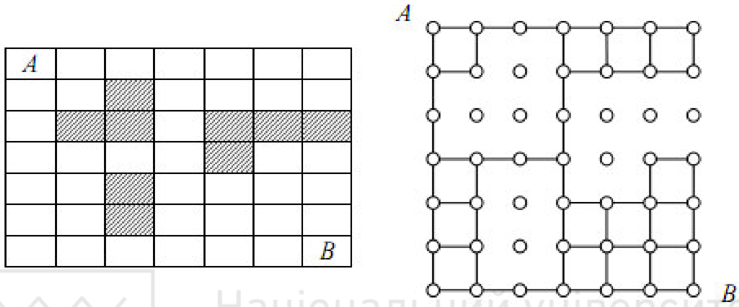
Виберемо будь-яку вершину із знайденої множини, наприклад, v_3. Визначимо далі множину

$$F_1(v_1) \cap D^{-1}(v_3) = \{v_4, v_5\} \cap \{v_4, v_5, v_6\} = \{v_4, v_5\}.$$

Виберемо будь-яку вершину із знайденої множини, наприклад, v_5. Тоді (v_1, v_5, v_3, v_6) – шуканий мінімальний шлях з v_1 до v_6 в орієнтованому графі G, а відстань між v_1 і v_6 дорівнює 3.

Очевидно, що хвильовий алгоритм може застосовуватися не тільки для орієнтованих, а й для неорієнтованих графів. В цьому випадку, пересування від однієї вершини до іншої можливе в обидві сторони.

Хвильовий алгоритм широко застосовується у розробці комп'ютерних ігор – коли необхідно визначити оптимальний маршрут пересування гравця з однієї точки віртуальної місцевості (карти) до іншої.



а) б)

Рис. 8.3. Карта місцевості та відповідний їй граф

Наведемо *приклад* такої задачі. Нехай потрібно знайти найкоротший маршрут з точки А до точки В на карті, яку зображено на рис. 8.3, а. На ній заштриховані комірки відповідають деяким перешкодам на шляху, тобто там гравець не зможе пройти. Також будемо вважати, що гравець може пересуватися лише вертикально чи горизонтально. Граф, що відповідає цій карті, зображено на рис. 8.3, б.

Внаслідок роботи хвильового алгоритму отримаємо такі індекси вершин – комірок карти (рис. 8.4, а). На рис. 8.4, б зображено два із знайдених маршрутів з вершини А у вершину В. Видно, що довжина знайденого маршруту дорівнює 12.

Можна зробити процес пошуку найкоротшого маршруту за допомогою хвильового алгоритму більш економним, а відтак, і більш швидким. Для цього треба розповсюджувати хвилю не тільки з початкової вершини А (перша хвиля), а й з кінцевої вершини В (друга хвиля). Для того, щоб відрізняти індекси першої та другої хвилі, індекси другої будемо позначати штрихом. Робота модифікованого алгоритму закінчується, коли обидві хвилі зустрінуться (рис. 8.5).

0	1	2	3	4	5	6
1	2		4	5	6	7
2			5			
3	4	5	6		10	11
4	5		7	8	9	10
5	6		8	9	10	11
6	7	8	9	10	11	12

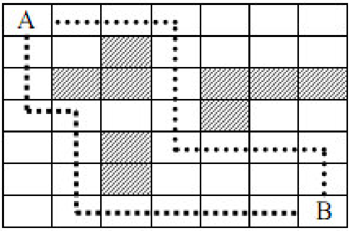


Рис. 8.4. Ілюстрація роботи хвильового алгоритму

Комірки, де дві хвилі зустрічаються, позначені подвійними лініями. З порівняння рис. 8.4 і рис. 8.5 видно, що знайдені найкоротші маршрути співпадають. І хоча в цьому прикладі економія склала лише одну комірку (яка не була відмічена), зрозуміло, що на більш складних (насичених «перешкодами») картах з великою кількістю комірок, робота модифікованого алгоритму буде більш ефективною, ніж простого хвильового алгоритму.

0	1	2	3	4	5	6
1	2		4	5	6	
2			5			
3	4	5	6/6'		4'	3'
4	5		5'	4'	3'	2'
5	6/6'		4'	3'	2'	1'
6/6'	5'	4'	3'	2'	1'	0'

Рис. 8.5. Двонаправлений хвильовий алгоритм

8.3. Знаходження найкоротшого шляху у мережі

Перейдемо до розгляду прикладних задач оптимізаційного аналізу зважених графів або мереж.

Нехай задано дорожну мережу, яка складається з дев'яти вершин (населених пунктів), з'єднаних ребрами (дорогами) так, як це показано на рис. 8.6. Необхідно знайти найкоротший шлях від вершини 1 до інших вершин мережі.

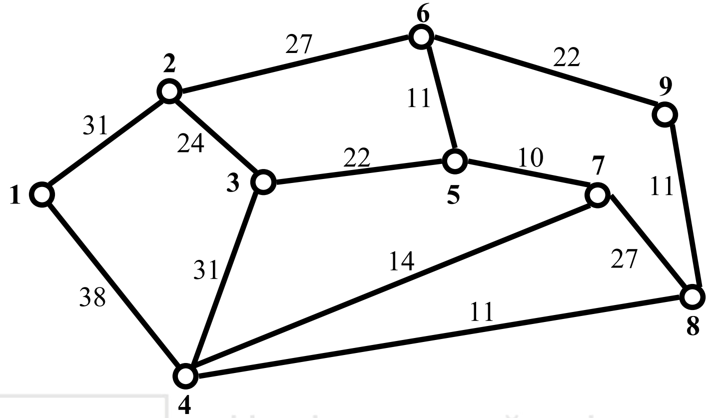


Рис. 8.6. Дорожна мережа

Вперше алгоритм визначення довжини найкоротшого шляху від фіксованої вершини до будь-якої іншої запропонував у 1952 р. нідерландський математик Е. Дейкстра. Цей алгоритм застосовний лише тоді, коли вага кожного ребра (дуги) графа додатна.

Застосуємо *алгоритм Дейкстри* для розв'язування поставленої задачі. В алгоритмі використовується три масиви розмірністю n, де n – кількість вершин заданого графа.

Перший масив a містить мітки з двома значеннями: 0 (вершину ще не розглянуто) і 1 (вершину уже розглянуто).

Другий масив b містить відстані – поточні найкоротші відстані від початкової вершини V_i до іншої вершини V_k.

Третій масив c містить номери вершин – k-ий елемент c_k є номером передостанньої вершини на поточному найкоротшому шляху з V_i до V_k. Матриця відстаней D_{ik} задає довжини ребер d_{ik}; якщо такого ребра немає, то d_{ik}

присвоюється значення дуже великого числа M, яке дорівнює «машинній нескінченності».

Алгоритм Дейкстри знаходження найкоротших шляхів у мережі складається з таких кроків.

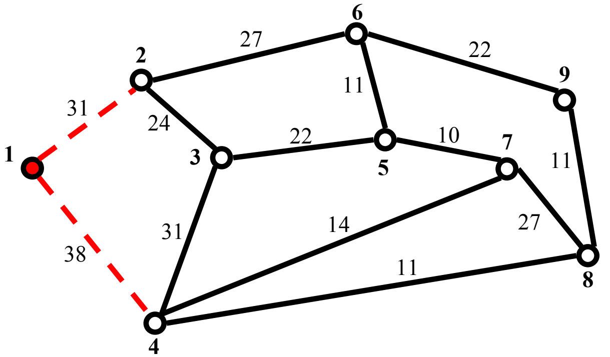
1. ***Ініціалізація***. Всім елементам масиву a, крім a_i (початок маршруту), присвоїти значення 0. $a_i := 1$. Всім елементам масиву b присвоїти значення відстаней з i-го рядка матриці D_{ik}. Всім елементам масиву c, крім c_i, присвоїти значення i (номер початкової вершини). $c_i := 0$.

2. ***Основна частина***. Знайти найменшу відстань b_j серед непомічених стовпців (тобто для тих j, для яких $a[j]=0$). Присвоїти $a[j]=1$ (помітити вершину). Розглянути всі маршрути, які проходять з вершини i через вершину j до вершини k. Обчислити їх довжини b_k. Якщо $b_k > b_j + d_{jk}$ (довжина нового маршруту менша від довжини старого), то присвоюємо $b_k := b_j + d_{jk}$; $c_k := j$. Якщо $b_k < b_j + d_{jk}$ (новий маршрут довший за старий) – ніяких змін у k-ий стовпець не вносимо. Основна частина повторюється до того часу, поки в масиві a залишається один нуль.

3. ***Результати***. Довжина найкоротшого шляху від вершини V_i до вершини V_k дорівнює b_k. Тепер потрібно визначити вершини, які входять в найкоротший шлях від V_i до V_k. Для цього служить масив c. Остання вершина – V_k, їй передує вершина з номером c_k, перед нею буде вершина, номер якої знаходиться у масиві c на місці c_k тощо.

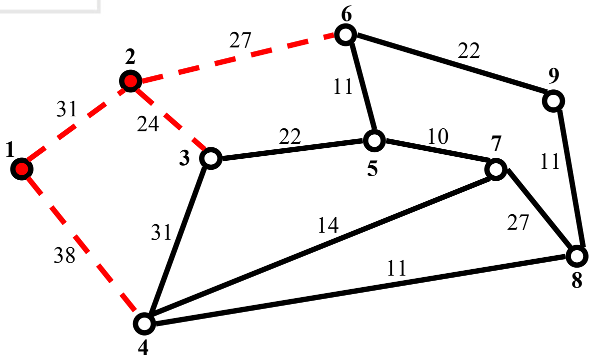
Процес розв’язування поставленої задачі за допомогою алгоритму Дейкстри проілюстровано наступною серією рисунків і таблиць.

1 крок



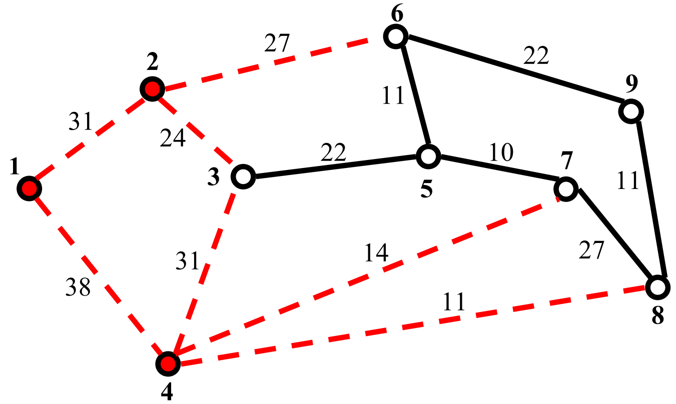
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	0	0	0	0	0	0	0	0	31
b	0	31	M	38	M	M	M	M	M	
c	0	1	1	1	1	1	1	1	1	

2 крок



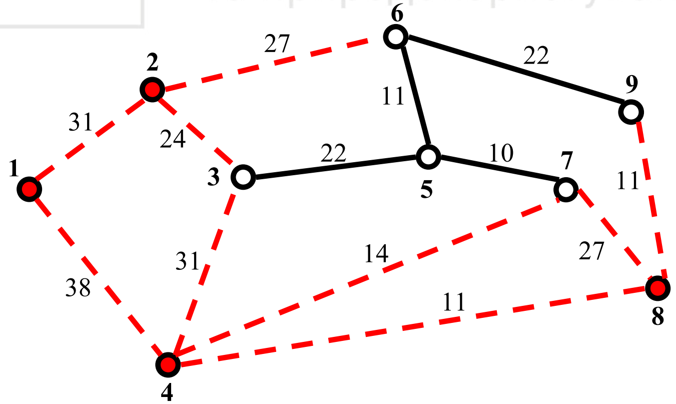
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	0	0	0	0	0	0	0	38
b	0	31	55	38	M	58	M	M	M	
c	0	1	2	1	1	2	1	1	1	

3 крок



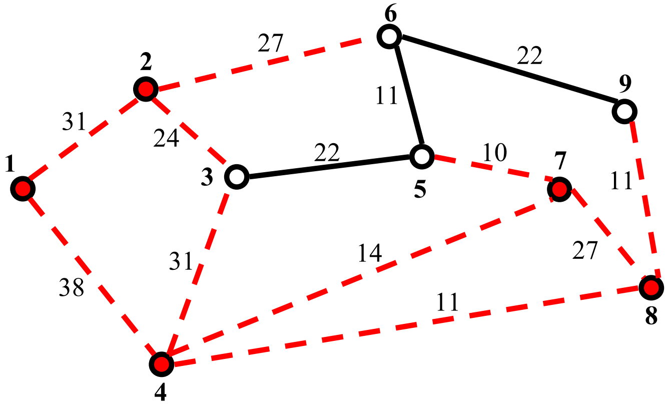
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	0	1	0	0	0	0	0	49
b	0	31	55	38	M	58	52	49	M	
c	0	1	2	1	1	2	4	4	1	

4 крок



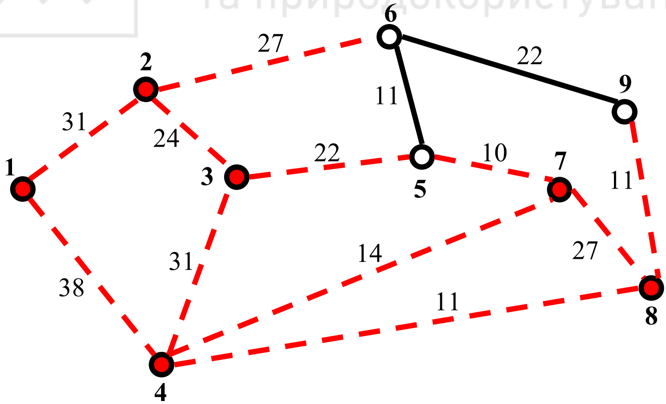
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	0	1	0	0	0	1	0	52
b	0	31	55	38	M	58	52	49	60	
c	0	1	2	1	1	2	4	4	8	

5 крок



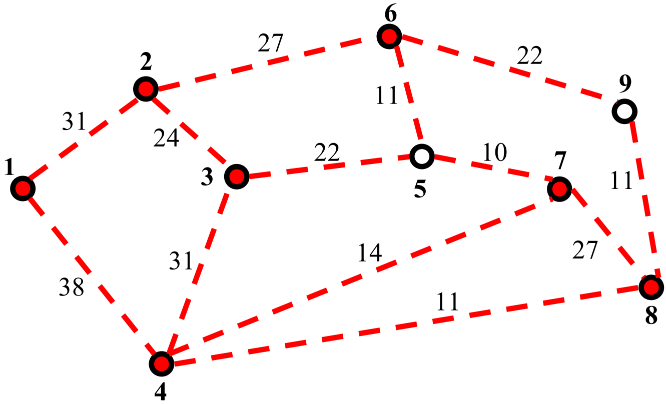
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	0	1	0	0	1	1	0	55
b	0	31	55	38	62	58	52	49	60	
c	0	1	2	1	7	2	4	4	8	

6 крок



	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	1	1	0	0	1	1	0	58
b	0	31	55	38	62	58	52	49	60	
c	0	1	2	1	7	2	4	4	8	

7 крок



	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$min\, b_k$
a	1	1	1	1	0	1	1	1	0	60
b	0	31	55	38	62	58	52	49	60	
c	0	1	2	1	7	2	4	4	8	

Таким чином, довжина найкоротшого шляху від вершини 1 до вершини 9 дорівнює 60. Щоб встановити сам цей шлях, необхідно скористатися масивом c. Остання вершина шляху – 9, їй передує вершина з номером 8, перед нею буде 4, перед нею – вершина 1. Отже, шуканий шлях 1 – 4 – 8 – 9.

Найкоротші відстані від пункту 1 до інших пунктів мережі визначаються за масивом b (шлях – за масивом c): з вершини 1 у вершину 2 – відстань 32 (шлях 1 – 2); з вершини 1 у вершину 3 – відстань 55 (шлях 1 – 2 – 3); з вершини 1 у вершину 4 – відстань 38 (шлях 1 – 4); з вершини 1 у вершину 5 – відстань 62 (шлях 1 – 4 – 7 – 5); з вершини 1 у вершину 6 – відстань 58 (шлях 1 – 2 – 6); з вершини 1 у вершину 7 – відстань 52 (шлях 1 – 4 – 7); з вершини 1 у вершину 8 – відстань 49 (шлях 1 – 4 – 8).

Для виконання алгоритму потрібно n раз переглянути масив b з n елементів, тобто алгоритм Дейкстри має квадратичний рівень складності.

8.4. Задача про оптимальну дорожну мережу

Розглянемо конкретну прикладну проблему. Є декілька міст, які необхідно з'єднати мережею доріг. Для кожної пари міст відома проектна вартість будівництва дороги, що їх з'єднує. Задача полягає в тому, щоб побудувати найдешевшу з можливих дорожніх мереж. Аналогічні задачі можна сформулювати для мереж ліній електропередач, водо- чи газопроводів тощо.

Якщо розглядати мережу доріг як граф, а вартість будівництва вважати пропорційною до довжини ребра (ваги ребра), що з'єднує відповідні вершини, то приходимо до задачі про побудову графа мінімальної довжини. Граф мінімальної довжини завжди є деревом, бо якби граф містив цикл, то можна було б видалити одне з ребер циклу, не порушивши зв'язності графа.

Отже, необхідно побудувати ***покриваюче дерево мінімальної довжини***. Для розв'язування задачі можна скористатись алгоритмом Прима–Краскала (так званим, ***жадібним алгоритмом***).

Алгоритм Прима–Краскала (1956 р.) для побудови мінімального покриваючого дерева описується наступним чином**.**

1. Серед ребер, які ще не розглядалися, обирається ребро мінімальної довжини.

2. Це ребро приєднується до раніше вибраних ребер за умови, що не утвориться цикл.

3. Після перегляду всіх ребер утворюється покриваюче дерево, яке і буде деревом мінімальної довжини.

Роботу алгоритму проілюструємо на *прикладі* (рис. 8.7). Загальна довжина покриваючого дерева становить 134 км.

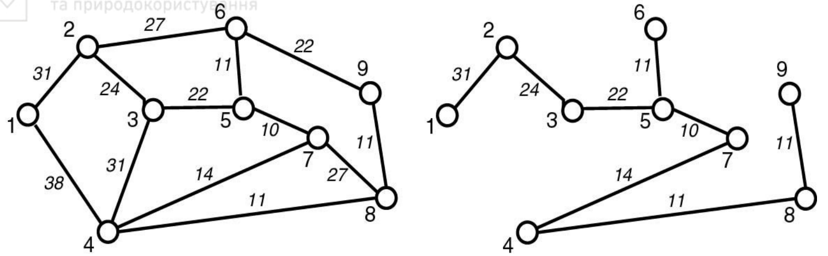


Рис. 8.7. Мінімальне покриття графа

Близькою до розглянутої вище є ***задача Р. Прима*** (1957 р.).

Нехай маємо N населених пунктів, заданих декартовими координатами на площині (x_i, y_i). Відстань між пунктами визначається за формулою

$$d_{ij} = \sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}\,. \tag{8.2}$$

Приймаємо спрощуючу умову, згідно з якою з'єднуючі дороги можуть бути прямими. Вартість будівництва дороги вважається пропорційною до її довжини d_{ij}. Необхідно побудувати дорожну мережу мінімальної вартості.

Алгоритм розв'язування має наступний вигляд.

1. Розраховується матриця відстаней за формулами (8.2).
2. За матрицею відстаней, використовуючи алгоритм Прима–Краскала, будується мінімальне покриваюче дерево.

Ілюстрацію розв'язку задачі Прима наведено на рис. 8.8.

Зауваження. Для побудови покриваючого дерева ***максимальної*** довжини алгоритми Дж. Краскала і Р. Прима є також застосовними.

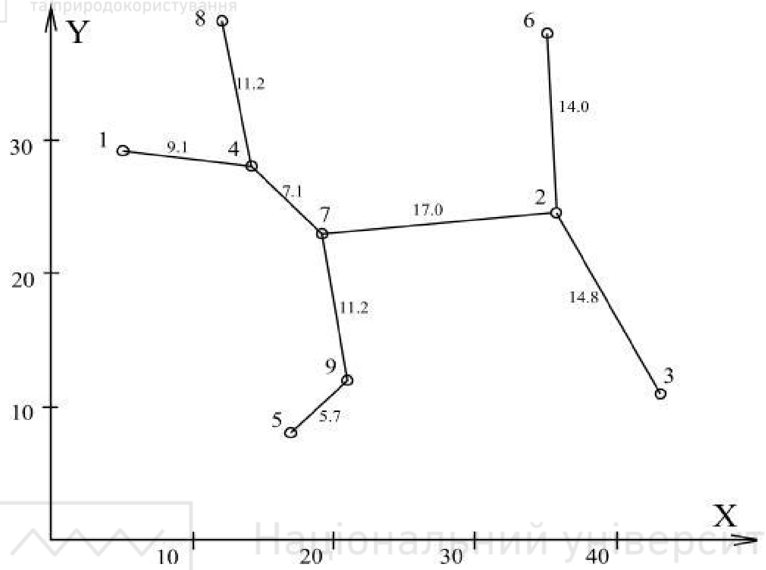


Рис. 8.8. Найкоротша дорожна мережа

8.5. Задачі про розміщення

Задача про розміщення школи

Одним з важливих практичних застосувань теорії графів є задачі про оптимальне розміщення соціальних об’єктів на дорожних мережах. У якості прикладу розглянемо дорожну мережу, яка з’єднує дев’ять населених пунктів (рис. 8.6). Кількість учнів у населених пунктах відома і становить: 20, 40, 60, 80, 10, 30, 50, 70, 90 відповідно. Необхідно оптимальним чином вибрати місце для розміщення нової школи.

В основу розв’язування задачі покладено міркування: сумарний шлях, пройдений всіма учнями дорогою до школи повинен бути мінімальним. Також доведено теорему [3], яка стверджує, що школа має бути розміщена у населеному пункті.

Перш за все необхідно побудувати матрицю найкоротших відстаней, тобто матрицю, яка містить довжини найкоротших ланцюгів, які зв’язують всі пункти мережі один з одним.

Для цього можна використати ***алгоритм Флойда–Воршала*** [3], в основу якого покладено операцію трикутника:

якщо $d[i,k]+d[k,j]<d[i,j]$, то $d[i,j]:=d[i,k]+d[k,j]$.

Ця операція заміняє маршрут $[i,\ j]$ на маршрут $[i,\ k,\ j]$, якщо він є коротшим.

Алгоритм Флойда–Воршала (основна частина)

```
for k := 1 to n do
    for i := 1 to n do
        for j := 1 to n do
            d [i,j] := min ( d [i,j], d [i,k] + d [k,j] );
```

Матриця відстаней $d[i,j]$ ініціалізується наступним чином. Якщо вершини i та j безпосередньо зв'язані, $d[i,j]$ дорівнює відстані між вершинами, інакше $d[i,j]$ дорівнює M (машинна нескінченність). Після завершення роботи алгоритму матриця $d[i,j]$ містить шукані значення довжини найкоротших ланцюгів.

У випадку невеликих мереж найкоротші маршрути між вершинами можна визначити вручну шляхом перебору варіантів.

В табл. 8.1 наведено *приклад* розрахунку матриці найкоротших маршрутів для мережі, представленої на рис. 8.6. До отриманої матриці відстаней приєднаємо ще один стовпчик, який містить кількості учнів p_j у населених пунктах.

В останньому рядку табл. 8.1 розміщено суми добутків відстаней відповідного стовпця на кількість учнів

$$f_i=\sum_{i=1}^{n} d_{ij}p_j\ . \qquad (8.3)$$

Таблиця 8.1

Задача про оптимальне розташування школи

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*к-сть учнів*
1	0	31	55	38	62	58	52	49	60	20
2	31	0	24	55	38	27	48	60	49	40
3	55	24	0	31	22	33	32	42	53	60
4	38	55	31	0	24	35	14	11	22	80
5	62	38	22	24	0	11	10	37	33	10
6	58	27	33	35	11	0	21	33	22	30
7	52	48	32	14	10	21	0	27	38	50
8	49	60	42	11	37	33	27	0	11	70
9	60	49	53	22	33	22	38	11	0	90
f_i	21370	18660	15060	**9560**	12390	12470	12040	10480	11760	

Найменша з цих сум (четвертий населений пункт) і вказує на оптимальне місце для розташування нової школи.

Оскільки у формулі (8.3) мінімізується сума, наведений алгоритм розв'язування задачі про розміщення школи називається ***мінісумним***.

Задача про розміщення пожежної частини

Іншим типом задач на розміщення є задачі про оптимальне розміщення пожежної частини, підрозділу охорони, станції швидкої допомоги. Головною метою при розв'язуванні таких задач є: дістатися до найбільш віддаленого пункту за мінімально можливий час.

Розглянемо наступну задачу. Необхідно оптимальним чином розмістити пожежну частину, яка виїжджає за викликом в один з пунктів мережі (рис. 8.6).

Є два варіанти розв'язування задачі: а) пожежна частина в населеному пункті; б) пожежна частина поза населеним пунктом. Зараз ми розглянемо лише більш простий перший варіант.

Таблиця 8.2

Задача про оптимальне розташування пожежної частини

	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***min***
1	0	31	55	38	62	58	52	49	60	62
2	31	0	24	55	38	27	48	60	49	60
3	55	24	0	31	22	33	32	42	53	55
4	38	55	31	0	24	35	14	11	22	55
5	62	38	22	24	0	11	10	37	33	62
6	58	27	33	35	11	0	21	33	22	58
7	52	48	32	14	10	21	0	27	38	**52**
8	49	60	42	11	37	33	27	0	11	60
9	60	49	53	22	33	22	38	11	0	60

Будуємо матрицю найкоротших маршрутів (табл. 8.2). До отриманої матриці відстаней приєднаємо ще один стовпчик, в якому виписуємо найбільше із значень поточного рядка (відстань до найбільш віддаленого пункту). У цьому (останньому) стовпчику знаходимо найменше значення. Це і буде оптимальний варіант розташування пожежної частини. Для нашого випадку це пункт 7.

Оскільки в розглянутому алгоритмі знаходиться мінімальне значення серед максимальних відстаней, метод розв'язування задачі про розміщення пожежної частини називається ***мінімаксним***.

Алгоритм розв'язування задачі у випадку, коли пожежна частина розміщується поза населеним пунктом, є значно складнішим.

Теми для роздумів і контрольні запитання

1. Що вивчає теорія графів?
2. Чим граф відрізняється від звичайної геометричної фігури?
3. Дайте означення графу. Що називають порядком графу?
4. Які є способи задання графів?
5. Поняття інцидентності і суміжності між вершинами і ребрами графу.
6. Як будується матриця інцидентності орграфу.
7. Що називається маршрутом, ланцюгом, циклом у графі?
8. Чим орграф відрізняється від звичайного графа?
9. Що таке зв'язний граф?
10. Який граф називається деревом?
11. Поняття покриваючого дерева графа.
12. Що називають мережею в теорії графів?
13. Степінь вершини графу, напівстепені входу і виходу.
14. Довжина маршруту (шляху) у графі.
15. Для чого використовується хвильовий алгоритм. Опишіть його.
16. Алгоритм Дейкстри визначення довжини найкоротшого шляху від фіксованої вершини графу до будь-якої іншої.
17. Алгоритм Прима–Краскала побудови покриваючого дерева мінімальної довжини.
18. Задачі про оптимальне розміщення соціальних об'єктів на дорожніх мережах. Алгоритм Флойда–Воршала.

9. ДОСЛІДЖЕННЯ СТІЙКОСТІ СИСТЕМ

9.1. Моделювання динамічних систем

Метою моделювання динамічних систем є вивчення детермінованої поведінки систем у часі. Дослідження динаміки поведінки систем дає змогу визначити перспективи їхнього розвитку, виявити можливі резерви, розробити комплекс адаптивних управлінських рішень, які забезпечать ефективне функціонування цих систем.

Модель популяції Мальтуса

Вперше моделювання динамічних систем як окремий розділ науки описав англійський вчений Р. Мальтус (1802 р.), який досліджував питання кількості живих організмів у деякій популяції (риби в морі, зайців та вовків у лісі тощо) з метою прогнозування зміни чисельності біологічних популяцій.

Нехай деяка популяція в момент часу t_0 має кількість осіб x_0. Як відомо, швидкість приросту біомаси пропорційна до наявної кількості біомаси. Якщо для розмноження популяції створені сприятливі умови (достатньо корму, немає хижаків), то динаміка популяції описується рівнянням Мальтуса (рис. 9.1)

$$\frac{dx}{dt} = rx\,, \tag{9.1}$$

розв'язком якого є експонента

$$x(t) = x_0 \cdot e^{r(t-t_0)}\,. \tag{9.2}$$

Моделювання динаміки продажів автомобілів на основі моделі популяцій

Рівняння Мальтуса може мати економічний зміст. Наприклад, позначимо x – кількість автомобілів, проданих за місяць; r – коефіцієнт, який залежить від кількості дилерів та якості їх роботи; $\dfrac{dx}{dt}$ – темп росту продажів. Якщо платоспроможність населення висока, ціна розумна, а попит на автомобілі незадоволений, то продажі будуть зростати. У такому випадку місячна кількість проданих авто буде описуватися рівнянням (9.1).

В економічній динаміці рівняння (9.1) називається ***рівнянням природного росту***. Цим рівнянням описуються також динаміка росту цін при постійному темпі інфляції.

Через певний час, настане насичення ринку і попит на автомобілі стабілізується. Рівняння, яке описує динаміку процесу місячних продажів автомобілів з врахуванням насичення ринку, має вигляд

$$\frac{dx}{dt} = rx - \frac{rx^2}{x_{max}}. \qquad (9.3)$$

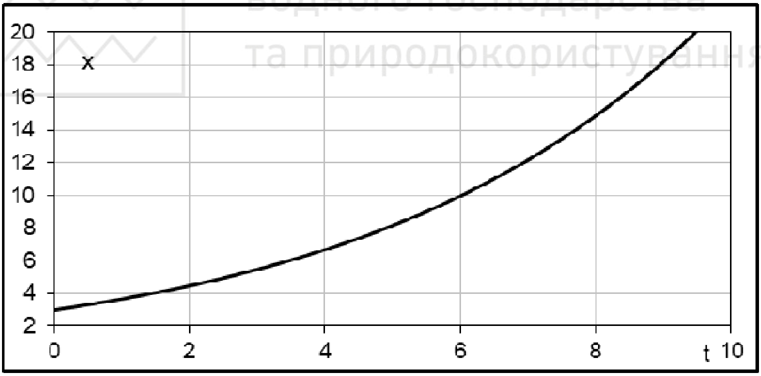


Рис. 9.1. Модель Мальтуса

Розв’язком рівняння (9.3) є логістична функція (рис. 9.2). Логістичне рівняння вперше запропонував Ферхюльст (1838 р.). Властивості логістичного рівняння:

1) при малій кількості автомобілів зростання місячних продажів відбувається за експоненціальним законом;

2) з часом розмір місячних продажів асимптотично наближається до деякого постійного числа x_{max}, яке визначається купівельною спроможністю населення.

Логістична крива (рис. 9.2), зокрема, описує динаміку поширення інформації, динаміку ефективності реклами тощо.

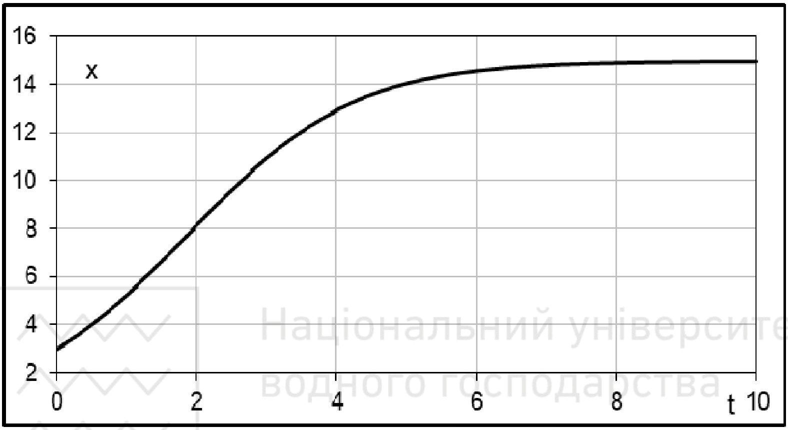


Рис. 9.2. Модель Ферхюльста

Розв’язавши рівняння

$$rx - \frac{rx^2}{x_{max}} = 0, \qquad (9.4)$$

можна визначити координати стаціонарних точок $x_1 = 0;\quad x_2 = x_{max}$, тобто таких точок, які відповідають стану рівноваги. Перша точка відповідає випадку, коли обсяги продажів нульові. Друга точка відповідає максимальним обсягам продажів.

Інколи динамічна система може мати декілька стаціонарних точок. Рис. 9.3 ілюструє можливі випадки економічного розвитку фірми:

а) $\frac{dx}{dt} < 0$ – випадок банкрутства;

б) $\frac{dx}{dt} = 0$ – одна стаціонарна точка;

в) $\frac{dx}{dt} = 0$ – дві стаціонарні точки.

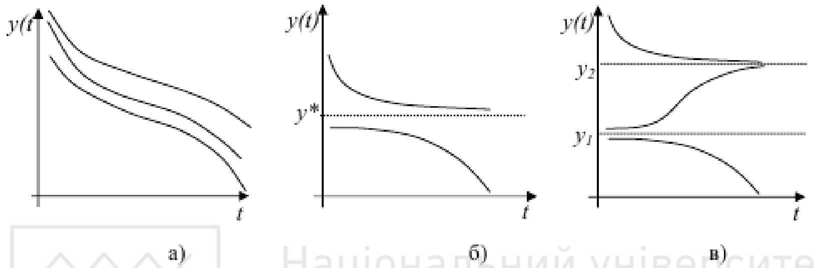


Рис. 9.3. а) випадок банкрутства; б) одна стаціонарна точка;
в) дві стаціонарних точки

Як видно з рис. 9.3, кінцева доля фірми залежить від початкової умови, тобто початкового рівня її доходу (прибутку).

Моделювання динаміки продажів автомобілів на основі моделі «хижак–жертва»

Розглянемо двохвидову модель «хижак–жертва», яка вперше була побудована італійським математиком Діно Вольтерра для пояснення коливань рибних уловів в Адріатичному морі, що носили циклічний характер. Суперництво хижака та жертви у цій моделі виражається зміною чисельності жертви за рахунок поїдання її хижаком, що в свою чергу впливає на чисельність хижака. В результаті кількість хижаків та жертв змінюються циклічно, тобто чергуються періоди зростання та спадання їхньої кількості.

У економічному сенсі (в термінах моделі продажу автомобілів) ця задача може бути сформульована наступним чином: автомобілі продаються у кредит за відсотковою ставкою

x_1, яка виступає регулятором купівельної активності населення ($\frac{dx_1}{dt}$ – темп росту кредитної ставки). Чим нижча кредитна (відсоткова) ставка, тим вища купівельна активність населення. Позначимо через x_2 – кількість автомобілів, проданих за місяць; $\frac{dx_2}{dt}$ – темп росту продажів автомобілів.

Необхідно дослідити зміну обсягу місячних продажів автомобілів у залежності від величини кредитної ставки. Місячні продажі будуть зростати доти, поки буде достатньо коштів у населення. Коли кошти населення почнуть вичерпуватися, кількість продажів почне зменшуватися. Тоді дилери з метою активізації покупок зменшать відсоткову ставку. Через деякий час, який необхідний для накопичення коштів населення, продажі знову почнуть рости. Після значного росту продажів відсоткова ставка x_1 знову збільшиться.

Отже, будуть спостерігатися циклічні коливання місячних продажів автомобілів у кредит та розміру кредитної ставки. Модель продажів буде мати вигляд

$$\begin{cases} \dfrac{dx_1}{dt} = -ax_1 + bx_1x_2, \\ \dfrac{dx_2}{dt} = cx_2 - dx_1x_2, \end{cases} \tag{9.5}$$

тут a, b, c, d – деякі додатні коефіцієнти.

Систему (9.5) називають моделлю «хижак–жертва» або системою Вольтерра-Лотки.

Коефіцієнт a визначає реакцію дилерів у випадку занадто великої відсоткової ставки; доданок $(-ax_1)$ – виражає зниження рівня кредитної ставки; коефіцієнт b визначає реакцію дилерів у випадку низької відсоткової ставки та активного попиту на автомобілі; доданок bx_1x_2 – виражає підвищення рівня кредитної ставки, яке залежить від попереднього рівня та обсягу місячних продажів. Коефіцієнт c – коефіцієнт приросту продажів, який залежить від ефективності роботи дилерів; доданок cx_2 описує збільшення місячних продажів при збільшенні кількості дилерських центрів; d – коефіцієнт зменшення продажів; доданок $\left(-dx_1x_2\right)$ описує зменшення продажів при великій кількості проданих автомобілів та великому розміру кредитної ставки.

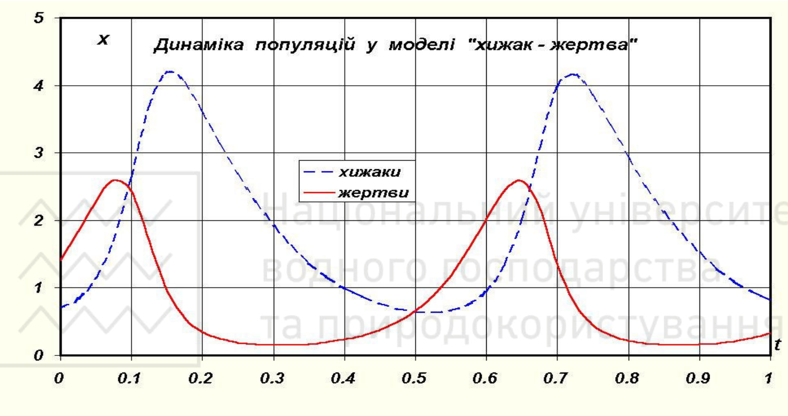


Рис. 9.4. Динаміка системи «хижак–жертва»

Результат розв’язування задачі «хижак–жертва» подано на рис. 9.4. Вид рисунка підтверджує циклічні коливання місячних продажів та розміру процентної ставки.

Якщо зобразити залежність $x_2\left(x_1\right)$ на графіку, то отримаємо замкнену траєкторію – ***фазовий портрет моделі*** (рис. 9.5). Стаціонарна точка процесу має наступні координати

$$x_{1c} = \frac{C}{D};\ x_{2c} = \frac{A}{B}.$$

Її зміст полягає у тому, що якщо вибирати стартові значення місячних продажів та процентної ставки якомога ближчими до координат стаціонарної точки, то фазова траєкторія буде вироджуватися в точку, тобто розміри місячних продажів і відсоткової ставки не змінюватимуться. Таким чином ця точка – це своєрідна точка ***динамічної рівноваги*** для системи продажів автомобілів у кредит. Однак у реальних умовах така рівновага неможлива через велику кількість зовнішніх стохастичних збурень, викликаних різними технологічними, економічними, юридичними та політичними чинниками.

Рис. 9.5. Фазовий портрет системи «хижак–жертва»

9.2. Формальний опис стану системи

Однією з найважливіших наукових проблем є розв’язання задачі прогнозування поведінки досліджуваної системи в часі та просторі на основі певних знань про її початковий стан. Ця задача зводиться до знаходження деякого закону, який дозволяє за наявною інформацією про систему, яка перебуває у початковий момент часу t_0 у точці x_0, визначити її майбутнє у будь-який момент часу $t > t_0$. Залежно від ступеня складності системи цей закон може бути детермінованим або ймовірнісним, може описувати зміну системи тільки в часі, тільки в просторі або ж просторово-часову еволюцію. В основу детермінованого опису динамічних систем покладено судження про те, що вся майбутня поведінка системи однозначно визначена її станом у початковий момент часу.

Станом (або вектором стану) системи називають сукупність N величин

$$x = \left(x_1, x_2, ..., x_N\right),$$

де N – розмірність опису.

Наприклад, для описання стану авторемонтного підприємства необхідно задати значення, що оцінюють наявні інструменти та системи діагностики x_1, трудові ресурси x_2, запаси запчастин x_3 тощо.

Величини x_i $\left(i = \overline{1, N}\right)$ називаються ***динамічними змінними***.

Вони визначають координати стану системи в k-вимірному ***просторі станів*** або у ***фазовому просторі***. При зміні стану системи змінюються її координати стану, тобто вони є змінними величинами, залежними від часу $x_i(t_1), x_i(t_2), \ldots, x_i(t_n)$.

Оператор еволюції Φ_t дозволяє за початковим станом $x(t_0)$ визначити стан системи в будь-який наступний момент часу $t_0 + t$

$$x(t_0 + t) = \Phi_t\left(x(t_0)\right).$$

Математично оператор еволюції може бути заданий диференціальними або різницевими рівняннями, дискретними відображеннями, матрицями, графами, марківськими ланцюгами тощо. Головною умовою є забезпечення однозначності прогнозу. В економіці найчастіше використовують моделі, описані за допомогою диференціальних і різницевих рівнянь.

Отже, ***динамічна система*** – це математична еволюційна модель, для опису якої необхідно:

а) визначити динамічні змінні та параметри системи;

б) визначити початковий стан системи $x(t_0)$;

в) задати оператор еволюції $\varPhi_t$, який дозволяє за початковим станом системи однозначно визначати майбутні стани*;*

Динамічні змінні – це величини x_i $\left(i=\overline{1,N}\right)$, що утворюють вектор стану $x=\left(x_1,x_2,...,x_N\right)$. Величини a_1, a_2, …, які у моделі системи є постійними чи можуть цілеспрямовано регулюватися, називають ***параметрами***. Параметри – це, здебільшого, коефіцієнти у рівняннях еволюції.

Зміну станів динамічної системи з часом називають ***процесом***. З погляду математики будь-який процес відповідає руху точки у ***фазовому просторі***. Найважливішою характеристикою фазового простору є його ***розмірність***, тобто мінімальна кількість параметрів, які необхідно задати для визначення стану системи. Розмірність фазового простору, яка забезпечує однозначне представлення всіх можливих станів системи, називається ***мінімальною розмірністю вкладення***. Розмірність вкладення є кількісною оцінкою складності системи.

Більшість економічних систем відносяться до класу відкритих нелінійних систем, які називають ***дисипативними***. Такі системи є нерівноважними через розсіювання ресурсів (гроші, енергія), отримуваних ззовні.

Кожному стану системи відповідає певна *точка фазового простору*. Процесу зміни станів відповідає ***траєкторія* –** лінія, яка з'єднує сусідні за часом точки фазового простору. Фазова траєкторія відображає поведінку системи під впливом деяких чинників. Множину усіх траєкторій називають ***фазовим портретом системи.*** Фазовий портрет являє собою сім'ю неперетинних кривих (рис. 9.5, 9.6). Це означає, що динамічна система може перебувати в кожному стані лише один раз.

В більшості дисипативних систем сукупність фазових траєкторій притягується до деякої скінченновимірної підмножини фазового простору, називаної ***атрактором*** (від англ. to attract – притягувати). Скінченновимірний атрактор визначає властивості усталеного з часом коливного процесу в системі. Коливання в економічних системах зумовлені коливаннями попиту та пропозиції, змінами відсоткових ставок та інших економічних умов виробництва.

Для хаотичних динамічних систем є характерним ***дивний атрактор***. Це означає, що невеликі зміни початкових умов системи можуть привести до кардинальних змін через певний проміжок часу (***ди́вний атра́ктор Ло́ренца***, рис. 9.6 – розв'язок системи Лоренца (1.4)). Інколи це ілюструють наступним прикладом Бенуа Мандельброта: змах крил метелика в одному місці земної кулі може викликати ураган в іншому її місці (Б. Мандельброт – «ефект метелика»).

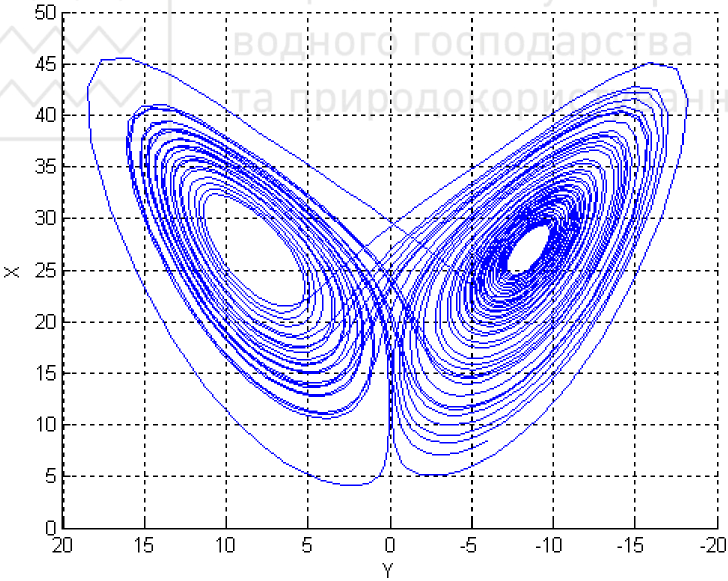


Рис. 9.6. Фазовий портрет системи Лоренца (двовимірна проекція)

Зміна параметрів стану може бути неперервною або ж дискретною. *Наприклад*: температура повітря чи зріст людини змінюються неперервно; ціна товару, кількість студентів у групі змінюється дискретно.

Позначимо через $x = x(t)$ – функцію входів системи, $y = y(t)$ – функцію виходів системи, $z = z(t)$ – функцію стану динамічної системи. Якщо всі ці функції неперервні, то поведінку динамічної системи описують за допомогою диференціальних рівнянь

$$\frac{dy}{dt} = f(t,x,z), \quad \frac{dz}{dt} = g(t,x,z). \tag{9.6}$$

Для дискретних систем зміна станів описується за допомогою різницевих рівнянь

$$z(t_{k+1}) - z(t_k) = g(t_k, z_k, x_k). \tag{9.7}$$

Розрізняють три характерних типи поведінки, або три режими, в яких може перебувати динамічна система: рівноважний, періодичний, перехідний.

Рівновага системи – це здатність її зберігати свій стан як завгодно довго (як за відсутності, так і за наявності зовнішніх впливів). Згідно із другим законом термодинаміки, при відсутності керуючих впливів складна ізольована система прямує до стану рівноваги, для якого є характерними дезорганізація та хаос.

Стан рівноваги може бути стійким, нестійким та байдужим. Під ***стійкістю*** системи розуміють здатність системи повертатися до стану рівноваги після виведення її з цього стану під впливом зовнішніх збурень. Стан рівноваги, до якого система здатна самостійно повертатися, називають ***стійким станом рівноваги*** (рис. 9.7, 2).

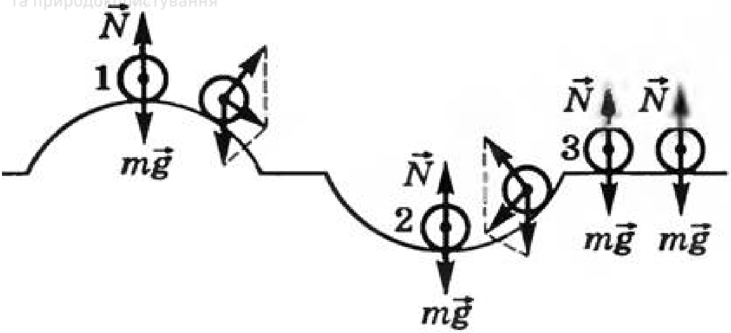


Рис. 9.7. Різні види механічної рівноваги
(1 – нестійка, 2 – стійка, 3 – байдужа)

При ***циклічному режимі функціонування*** система через деякі проміжки часу приходить до одного і того ж стану (потрапляє в одну і ту ж точку фазового простору). Циклічність є одним з видів стійкості. *Приклади*: чергування пір року, економічні цикли.

Перехідним режимом називається рух динамічної системи з одного стійкого режиму до іншого. Швидкість перехідного процесу характеризує інерційність системи.

9.3. Дослідження стійкості динамічних систем

В теорії систем при розв'язанні різних задач поняття стійкості систем формулюється по-різному. Це може бути поняття про математичну стійкість задачі при зміні початкових даних, про структурну стійкість соціально-економічних систем тощо. Але найбільш вживаним у теорії систем є строго математично сформульоване поняття стійкості за Ляпуновим.

Визначення стійкості. Динамічна система називається ***автономною***, якщо час t явно не входить у модель системи. Розглянемо випадок двох змінних. Нехай в момент часу t_0 система знаходиться у точці з координатами (x_0, y_0). З плином часу ця точка рухається, породжуючи фазову траєкторію $T_0(x, y)$. Розглянемо іншу фазову траєкторію цієї ж системи

$T_1(x,y)$, яка в момент часу t_0 стартувала від точки з координатами (x_1, y_1), розташованої дуже близько (в околі δ) точки (x_0, y_0). Якщо в довільний момент часу t відстань між траєкторіями $T_0(x,y)$ і $T_1(x,y)$ буде малою (не перевищить заданої величини ε), то траєкторія $T_0(x,y)$ називається **стійкою за Ляпуновим** (рис. 9.8).

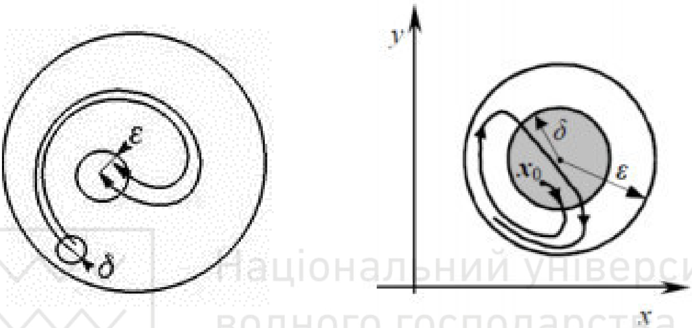


Рис. 9.8. Приклади руху, стійкого за Ляпуновим

Асимптотична стійкість є більш сильним поняттям ніж стійкість за Ляпуновим. Точка x_0 називається асимптотично стійкою, якщо вона стійка за Ляпуновим і, крім того, з плином часу обидві спостережувані траєкторії постійно зближуються (рис. 9.9).

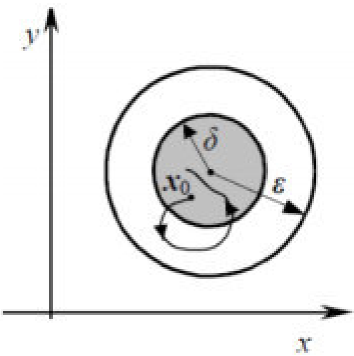


Рис. 9.9. Приклад руху, асимптотично стійкого за Ляпуновим

Точка x_0 називається ***нестійкою за Ляпуновим***, якщо не можна вказати таку область допустимих відхилень від неї (область ε), для якої жоден рух, що розпочинається всередині околу δ, ніколи не досягне межі області ε (рис. 9.10).

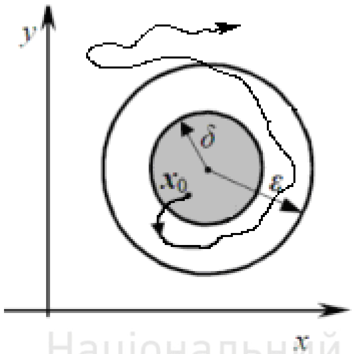


Рис. 9.10. Приклад руху, нестійкого за Ляпуновим

Алгоритм дослідження динамічної системи на стійкість. Для практичних задач корисно мати просту схему дослідження на стійкість. Ця схема ґрунтується на лінійному наближенні рівнянь математичної моделі системи (нелінійні доданки відкидаються). Наведемо її для випадку систем розмірності 2.

Нехай дано систему двох диференціальних рівнянь, які описують динаміку деякої двовимірної системи

$$\begin{cases} \dfrac{dx}{dt} = a(x,y); \\ \dfrac{dy}{dt} = b(x,y), \end{cases} \tag{9.8}$$

де $a(x,y);\quad b(x,y)$ – задані нелінійні функції.

Перший етап – відшукання *стаціонарних режимів*. Стаціонарні точки (9.8) визначаються з системи двох рівнянь з двома невідомими

$$a(x,y)=0;\ b(x,y)=0. \tag{9.9}$$

Ця система може мати багато розв'язків, частина з яких може бути комплексними. Кожен розв'язок описує стаціонарну точку, яку необхідно дослідити на стійкість.

Другий етап – *лінеаризація системи*. Вибираємо конкретну стаціонарну точку (x_0, y_0) таку, що для неї виконуються умови $a(x_0, y_0)=0;\ \ b(x_0, y_0)=0$ і будуємо для неї лінеаризовану систему. Для цього необхідно обчислити матрицю частинних похідних цієї системи в точці (x_0, y_0)

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}=\begin{pmatrix} \dfrac{\partial a}{\partial x} & \dfrac{\partial a}{\partial y} \\ \dfrac{\partial b}{\partial x} & \dfrac{\partial b}{\partial y} \end{pmatrix}_{\substack{x=x_0 \\ y=y_0}}. \tag{9.10}$$

Третій етап – *знаходження власних значень лінеаризованої системи*. Будуємо визначник лінеаризованої матриці і записуємо характеристичне рівняння

$$\begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix}=0.$$

Розкривши визначник у характеристичному рівнянні, отримаємо квадратне рівняння для знаходження власних значень λ

$$\lambda^2 - S\lambda + D = 0, \tag{9.11}$$

де $S = a_{11} + a_{22}$ – слід матриці A (сума діагональних елементів), $D = a_{11}a_{22} - a_{12}a_{21}$ – її визначник.

З формули Вієта відомо, що для коренів рівняння (9.11) λ_1 і λ_2 виконуються умови

$$S = \lambda_1 + \lambda_2 \text{ і } D = \lambda_1 \lambda_2 . \qquad (9.12)$$

Ляпунов показав, що система є стійкою при виконанні наступних умов

$$\lambda_1 < 0; \lambda_2 < 0 \text{ або } Re\lambda_1 < 0; Re\lambda_2 < 0 , \qquad (9.13)$$

де $\mathrm{Re}\lambda$ – дійсна частина комплексного числа λ .

Якщо $\lambda_1 > 0$, або $\lambda_2 > 0$, або $\mathrm{Re}\lambda_1 > 0$, або $\mathrm{Re}\lambda_2 > 0$, то система є нестійкою. Якщо $\lambda_1 = 0; \lambda_2 = 0$ (або $Re\lambda_1 = 0; Re\lambda_2 = 0$), то метод лінійного наближення не може дати відповідь на питання про стійкість системи.

Отже, враховуючи рівності (9.12) і (9.13) отримуємо, що динамічна система буде стійкою лише за умови

$$S < 0 ,\ D > 0 .$$

Приклад. Нехай нелінійна динамічна система задана системою диференціальних рівнянь

$$\begin{cases} \dfrac{dx_1}{dt} = x_1 - 3x_2; \\ \dfrac{dx_2}{dt} = x_1 + x_2 + x_1^2. \end{cases} \qquad (9.14)$$

Для знаходження координат стаціонарної точки складаємо систему рівнянь

$$\begin{cases} x_1 - 3x_2 = 0; \\ x_1 + x_2 + x_1^2 = 0. \end{cases} \tag{9.15}$$

Система (9.15) має два розв’язки: $x_1 = 0; x_2 = 0$ та $x_1 = -\dfrac{4}{3}; x_2 = -\dfrac{8}{3}$, які описують стаціонарні точки системи (9.14). Дослідимо першу стаціонарну точку.

Лінеаризована матриця частинних похідних має наступний вигляд

$$\begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}$$

Відповідно, характеристичне рівняння системи має вигляд

$$\begin{vmatrix} 1-\lambda & -3 \\ 1 & 1-\lambda \end{vmatrix} = 0, \tag{9.16}$$

або

$$\lambda^2 - 2\lambda + 4 = 0. \tag{9.17}$$

Детермінант рівняння (9.17)

$$D = b^2 - 4ac = 4 - 4 \cdot 1 \cdot 4 = -12.$$

Отже, корені рівняння (9.17) комплексні і мають наступний вигляд: $\lambda_1 = 1 + i\sqrt{3}; \lambda_2 = 1 - i\sqrt{3}$. Оскільки дійсна частина коренів $Re(\lambda_{1,2}) = 1 > 0$, то досліджувана стаціонарна точка відповідає нестійкому стану системи. Це означає, що всі фазові траєкторії, які знаходяться в околі цієї точки, з плином часу будуть віддалятися (відштовхуватися) від неї.

Дослідження стійкості моделі «хижак–жертва»

Дослідимо на стійкість розглянутим вище аналітичним методом систему динаміки продажів автомобілів, змодельовану у п. 9.1 на основі системи «хижак–жертва».

При дослідженні використовуємо алгоритм, описаний вище.

Спочатку знайдемо стаціонарні точки. Для цього слід розв’язати систему рівнянь

$$\begin{cases} ax - bxy = 0, \\ -cy + dxy = 0. \end{cases} \tag{9.18}$$

Система (9.18) має два розв’язки. Перший з них $x_0 = 0, \quad y_0 = 0$ відповідає ситуації, коли продажі відсутні.

Більш цікавим є другий розв’язок

$$y_0 = \frac{a}{b}, \quad x_0 = \frac{c}{d}. \tag{9.19}$$

Проведемо дослідження стаціонарної точки (9.19) на стійкість. Вирази для похідних будуть наступними:

$$\frac{\partial \dot{x}}{\partial x} = a - by; \quad \frac{\partial \dot{x}}{\partial y} = -bx; \quad \frac{\partial \dot{y}}{\partial x} = dy; \quad \frac{\partial \dot{y}}{\partial y} = -c + dx .$$

Матриця лінеаризованої системи має вигляд

$$A = \begin{pmatrix} a - by & -bx \\ dy & -c + dx \end{pmatrix}. \tag{9.20}$$

Підставивши координати стаціонарної точки (9.19) у (9.20) отримаємо матрицю

$$A = \begin{pmatrix} 0 & -\frac{bc}{d} \\ \frac{ad}{b} & 0 \end{pmatrix}.$$

Характеристичне рівняння має вигляд

$$\begin{vmatrix} 0-\lambda & -\frac{bc}{d} \\ \frac{ad}{b} & 0-\lambda \end{vmatrix} = 0, \tag{9.21}$$

або

$$\lambda^2 + ac = 0. \tag{9.22}$$

Розв'язками рівняння (9.22) є $\lambda_1 = i\sqrt{ac};\ \lambda_2 = -i\sqrt{ac}$. $Re\,\lambda_1 = 0;\ Re\,\lambda_2 = 0$, отже лінійне наближення не може дати відповідь про стійкість системи. Тому, тут потрібно проводити дослідження на стійкість іншими методами.

Проведений аналіз свідчить про те, що поведінка розглянутої системи не є стійкою: малі зміни параметрів моделі (зокрема, врахування обмеженості ресурсів) може привести до якісної зміни поведінки. Наприклад, рівноважний стан може стати стійким, і коливання продажів будуть затухати, що приведе до їх стабілізації. Можлива і протилежна ситуація, коли мале відхилення від положення рівноваги через деякий час призведе до повної зупинки продажів. На питання про те, який з цих сценаріїв реалізується, модель Вольтерра–Лотки «хижак–жертва» (9.5) не може дати відповіді.

Для стабілізації цієї моделі необхідно змінити її праву частину. Модифікація правої частини полягає у врахуванні ефектів насичення продажів та обмеженості купівельної спроможності покупців.

9.4. Класифікація точок рівноваги

Розглянемо характер стійкості особливих точок системи двох диференціальних рівнянь для різних випадків коренів характеристичного рівняння (9.11), що відповідає матриці лінеаризованої системи (9.10). Зауважимо, що ці випадки поширюються і на системи диференціальних рівнянь більш високих порядків.

Нехай λ_1 і λ_2 – корені характеристичного рівняння (9.11), які, як уже зазначалося, визначають стійкість системи у даній стаціонарній точці. Можливі наступні випадки.

Випадок1. λ_1 і λ_2 – *дійсні* ($\lambda_1, \lambda_2 \in R_1$).

1.1. Якщо λ_1 і λ_2 різних знаків, то фазовий портрет називається ***сідлом*** (рис. 9.11). Дві пари траєкторій проходять через стаціонарну точку, а інші виглядають так, як горизонталі на карті місцевості, котра є гірським перевалом. Динаміка нестійка.

1.2. Якщо $\lambda_1 > 0$ і $\lambda_2 > 0$, то фазовий портрет називається ***нестійким вузлом***. Кожна фазова траєкторія «примикає» до особливої точки і з часом віддаляється від неї (рис. 9.12). Динаміка нестійка.

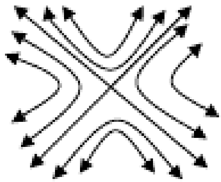


Рис. 9.11. Фазовий портрет «Сідло»

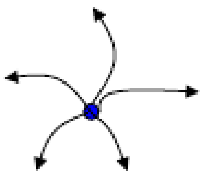


Рис. 9.12. Фазовий портрет «Нестійкий вузол»

1.3. Якщо $\lambda_1 < 0$ і $\lambda_2 < 0$, то фазовий портрет називається ***стійким вузлом***. Усі траєкторії проходять через стаціонарну точку і з часом до неї наближаються (рис. 9.13). Динаміка стійка.

Випадок 2. λ_1 і λ_2 – *комплексно–спряжені* ($\lambda_{1,2} = \alpha \pm i\beta$).

2.1. Якщо $a < 0$, то фазовий портрет називається ***стійким фокусом***. Траєкторії асимптотично наближаються до особливої точки, «намотуючись» на неї у вигляді спіралей (рис. 9.14). Динаміка стійка.

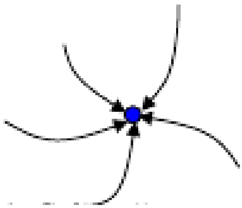


Рис. 9.13. Фазовий портрет «Стійкий вузол»

Рис. 9.14. Фазовий портрет «Стійкий фокус»

2.2. Якщо $\alpha > 0$ (дійсна частина додатна), то фазовий портрет називається ***нестійким фокусом*** (рис. 9.15). Динаміка нестійка.

2.3. Якщо $\alpha = 0$, тобто характеристичні числа $\lambda_{1,2} = \pm i\beta$ мають тільки уявну частину, то фазовий портрет називається ***граничним циклом*** (рис. 9.16). Цей тип поведінки відповідає описаній вище моделі «хижак–жертва». Висновок щодо типу динаміки зробити неможливо.

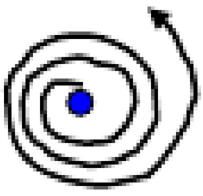


Рис. 9.15. Фазовий портрет «Нестійкий фокус»

Рис. 9.16. Фазовий портрет «Граничний цикл»

Класифікація типів поведінки фазових кривих в околі особливої точки була здійснена французьким математиком Анрі Пуанкаре.

9.5. Павутиноподібна модель

Цікавою ілюстрацією поведінки економічної системи, що відображає характер стійкості особливих точок, може бути ***динамічна павутиноподібна модель ринкової рівноваги (модель Курно)***. Назву павутиноподібна модель одержала завдяки геометричній інтерпретації − на площині $q0p$ (p – ціна, q – пропозиція) відповідний ітераційний процес розвитку системи подається у вигляді «павутини», що «намотується» на криві попиту та пропозиції.

Павутиноподібна модель дозволяє досліджувати стійкість цін і обсягів товарів на ринку, що описується традиційними кривими попиту та пропозиції за наявності запізнювання в часі (лагу) одного з процесів. Пропозиція S реагує на зміну ціни p з лагом в один період, тоді, як попит D визначається ціною, і обидві ці залежності лінійні

$$\begin{cases} D_t = a + bp_t; \\ S_t = a_1 + b_1 p_{t-1}. \end{cases} \tag{9.23}$$

Будемо вважати, що ціна, встановлена на початку виробничого циклу, не змінюється протягом цього циклу. Пропозиція визначається цією ціною. Нова пропозиція з'явиться на ринку лише після закінчення циклу. Крім того, будемо вважати, що у кожному періоді попит поглинає весь обсяг пропозиції (ненасичений ринок). Балансове рівняння моделі має вигляд

$$D_t = S_t\,. \tag{9.24}$$

З урахуванням (9.24) модель (9.23) набуває вигляду

$$bp_t - b_1 p_{t-1} = a_1 - a . \qquad (9.25)$$

Рівноважний розв'язок рівняння (9.25) має вигляд

$$p^* = \frac{a_1 - a}{b - b_1} . \qquad (9.26)$$

Цей розв'язок відповідає рівноважному значенню ціни.

Розглянута вище модель описує ***економічний закон попиту та пропозиції***, згідно з яким:

• кількість товару, що його можна продати на ринку (тобто попит), змінюється у напрямку, протилежному до зміни ціни товару;

• кількість товару, який виробляють і доставляють на ринок (тобто пропозиція), змінюється у тому самому напрямку, що й ціна;

• водночас реальна ринкова ціна формується на рівні, на якому попит і пропозиція наближено дорівнюють одне одному (приблизно збігаються з деякою заданою точністю), тобто перебувають у рівновазі.

Умова стабільності моделі полягає в тому, що ріст ціни приводить до більшого розширення пропозиції у порівнянні з попитом, а падіння ціни – до більшого розширення попиту у порівнянні з пропозицією.

Графік попиту D_t на площині $q0p$ має від'ємний кутовий коефіцієнт відносно осі $0p$: $b < 0$, а пропозиція S_t – додатний кутовий коефіцієнт відносно осі $0p$: $b_1 < 0$. Таким чином, $\frac{b_1}{b} < 0$ і ціна здійснюватиме рухи навколо свого рівноважного значення. Амплітуда цих періодичних коливань може бути зростаючою, постійною або спадаючою (рис. 9.17) в залежності від співвідношення між $|b_1|$ і $|b|$.

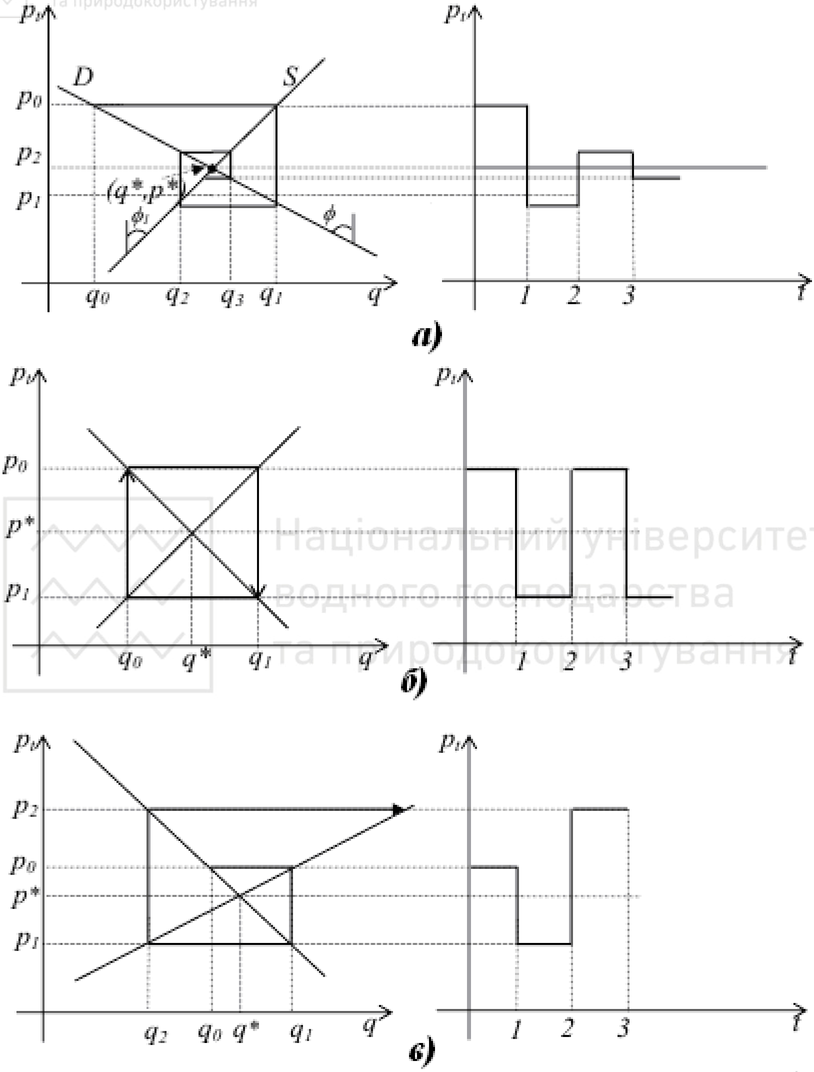


Рис. 9.17. Ілюстрація до теореми про ринкову рівновагу у павутиноподібній моделі: а) рівновага стійка; б) регулярні коливання; в) рівновага нестійка

Теорема (Курно) про ринкову рівновагу у павутиноподібній моделі:

- якщо $|b_1| > |b|$, то пропозиція зростає швидше від попиту і це приводить до дестабілізації ринку (нестійка рівновага) (рис. 9.17, в);
- якщо $|b_1| < |b|$, то пропозиція зростає повільніше від попиту і це приводить до стабілізації ринку (стійка рівновага) (рис. 9.17, а);
- якщо $|b_1| = |b|$ пропозиція дорівнює попиту і це приводить до циклічних коливань ринку (рис. 9.17, б).

Теми для роздумів і контрольні запитання

1. Поняття динамічної системи. Моделювання динамічних систем.
2. Яка мета моделювання динамічних систем?
3. Що таке детермінована динамічна система?
4. Модель динаміки популяції (модель Мальтуса). Припущення моделі.
5. Наведіть приклади систем, динаміку яких можна змоделювати на основі моделі популяції Мальтуса.
6. Логістичне рівняння. Властивості логістичної функції. Динаміку яких процесів можна описати логістичним рівнянням? Наведіть приклади.
7. Модель Вольтерра «хижак–жертва». Припущення моделі.
8. Який фізичний (економічний) сенс мають параметри системи Вольтерра–Лотки (моделі «хижак–жертва»)?
9. Як визначається стаціонарна точка процесу, що описується моделлю «хижак–жертва», і який її зміст?
10. Що таке фазовий портрет моделі?
11. Точка динамічної рівноваги системи.
12. Вектором стану системи називають…
13. Що таке динамічні змінні? Що визначають динамічні змінні?
14. Яка різниця між динамічними змінними і параметрами системи?

15. Поняття фазового простору (або простору станів) динамічної системи.
16. Що таке розмірність фазового простору? Що визначає розмірність фазового простору динамічної системи?
17. Що таке фазова траєторія? Що визначає фазова траєторія динамічної системи?
18. Чи можуть фазові траєторії, що утворюють фазовий портрет динамічної системи, перетинатися між собою? Відповідь обґрунтуйте.
19. Що називається оператором еволюції динамічної системи?
20. Як математично може бути заданий оператор еволюції?
21. Що визначає оператор еволюції динамічної системи?
22. Формальний опис стану динамічної системи включає в себе…
23. Що таке атрактор (дивний атрактор)?
24. Рівновага та стійкість системи.
25. Визначення стійкості системи за Ляпуновим. Яка система називається стійкою, асимптотично стійкою і нестійкою?
26. Схема дослідження на стійкість двовимірної системи. Умови стійкості.
27. Класифікація точок рівноваги двовимірної системи за коренями характеристичного рівняння лінеаризованої системи.
28. Динамічна павутиноподібна модель ринкової рівноваги (модель Курно). Теорема Курно про ринкову рівновагу у павутиноподібній моделі.

10. МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ СИСТЕМ МЕТОДАМИ НЕЛІНІЙНОЇ ДИНАМІКИ

10.1. Ідентифікація систем

Результати діяльності нелінійних динамічних процесів, так само як і лінійних, відображаються у вигляді часових рядів їх параметрів. Проте, такі часові ряди мають складну структуру і, окрім тренду, містять різні види коливань, зокрема, сезонні та циклічні.

Структуру таких рядів, як правило, не можна описати за допомогою відомих функцій, оскільки для різних ділянок часового ряду набір цих функцій буде різним, тобто в цьому разі можна говорити про ряди зі змінною структурою, які характерні для нелінійних динамічних процесів.

Для прикладу, такі ряди спостерігаються в динаміці цін на ринках капіталу.

Лише в останні роки завдяки розвитку математичних методів нелінійної динаміки та комп'ютерних технологій з'явилася можливість досліджувати такі процеси. Взагалі кажучи, в певному сенсі, будь-який динамічний процес зрештою є детермінованим, і моделювання його як реалізації випадкового процесу є зручним спрощенням. Невипадковий часовий ряд відображає невипадкову природу впливів. Флуктуації даних відображають флуктуації впливаючих чинників і відображають властиву їм кореляцію.

Детерміновані процеси, що виглядають як випадкові, у теорії нелінійностей називають ***детермінованим хаосом***. Добре відомо, що просте детерміноване нелінійне різницеве рівняння може породжувати надзвичайно складні часові траєкторії, які видаються випадковими.

Наприклад, рівняння, яке описує динаміку ціни облігацій y_t на фінансовому ринку

$$y_{t+1} = \alpha y_t (1 - y_t). \qquad (10.1)$$

В економічних застосуваннях значення параметру α лежить між 1 та 4. Таким чином виключають від'ємні значення рівноваги для y_t і уникають прямування процесу до нескінченності. Проте за зміни α від 1 до 4 динаміка системи зазнає суттєвих змін. Зокрема, для $1<\alpha<3$ за будь-яким відхиленням від $y_t=0$, динаміка процесу прямує до рівноважного значення $y_t = 1 - \frac{1}{\alpha}$, а для $3{,}75<\alpha<4$ спостерігатиметься нескінченна кількість циклів з різною періодичністю і нескінченне число положень рівноваги з еволюцією процесу залежно від початкового його стану. Такий тип поведінки називають «динамічний хаос». Властивістю цього процесу є те, що хоча він детермінований, випадкове блукання є задовільною моделлю для описання механізму породження даних. У цьому випадку зміни y_t неможливо передбачити, хоча всю траєкторію перебігу процесу цілком можна передбачити [31, С. 37–38].

Однією з перших класичних моделей, якими намагались описати змінне середнє була кусково-лінійна модель. У ній припускалося, що весь період спостережень може бути розбитий на певні проміжки часу, спостереження на яких описуються лінійними регресійними моделями. Для покращення лінійної апроксимації таких моделей до реальних процесів зараз розроблено низку, так званих, порогових моделей. Узагальнення тут відбулося як у напрямку відмови від простої лінійності, так і в більш гнучкому механізмі переключення від одного режиму до іншого. Але, у багатьох прикладних задачах порогові моделі не можуть бути застосовані, оскільки умовне математичне сподівання не має необхідної гладкості на «порогах». Неперіодичне переключення між різними станами в багатьох економічних, і не тільки, явищах спонукало до виникнення і вивчення класу моделей з переключеннями, наприклад, модель з марківськими переключеннями.

Досліджуючи специфіку часових рядів можна зробити висновки щодо характеру породжуючих їх систем. Зокрема, на основі дослідження часових рядів визначають розмірність системи, можливості її прогнозування та максимального горизонту прогнозування, характер динаміки системи.

Обов’язковим етапом, який передує моделюванню та прогнозуванню системи, є передпрогнозний аналіз часових рядів, який називають ***ідентифікацією системи***. На етапі ідентифікації, як правило, визначають ступінь випадковості системи, досліджують автокореляції тощо.

Визначення ступеня випадковості системи

Згідно з критерієм Кендала, якщо фактична кількість поворотних точок P_f часового ряду: $P_f < P_L$, то це означає, що для ряду характерні трендові ділянки. Такий ряд, як уже зазначалося, називають ***персистентним (трендостійким)***. Для прогнозування таких рядів доцільно використовувати трендові моделі.

Якщо $P_f > P_U$, це означає надмірну кількість коливань, викликану деяким зовнішнім чинником. Такий ряд називають ***антиперсистентним (реверсивним)***. Для моделювання і прогнозування таких рядів використовують гармонічні моделі та авторегресійні моделі (***авторегресія*** відображає часткову повторюваність окремих ділянок ряду).

У разі $P_L \le P_f \le P_U$ ряд вважають ***випадковим*** (стохастичним). Стохастичні ряди моделювати найважче. Для цього використовують теорію випадкових процесів та комп’ютерне імітаційне моделювання.

Дослідження автокореляцій

Потужним інструментом дослідження структури часового ряду є кореляційний аналіз, використовуючи який можна отримати відповіді на наступні питання: чи має ряд тренд? чи присутня циклічна компонента ряду? Кореляційну залежність між послідовними рівнями часового ряду називають автокореляцією рівнів ряду. Для дослідження такої залежності необхідно побудувати ряд перших різниць

$$e_t = x_{t+1} - x_t \,. \tag{10.2}$$

Коефіцієнт автокореляції першого порядку має вигляд

$$\rho_1 = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sqrt{\sum_{t=2}^{n} e_t^2 \cdot \sum_{t=2}^{n} e_{t-1}^2}} \,. \tag{10.3}$$

Коефіцієнт ρ_1 вимірює залежність між сусідніми рівнями ряду в моменти часу t і $t-1$, тобто для лагу 1. ***Лагом*** L (з нім. *der Lagoperator*) називають різницю в часі між рівнями одного і того ж часового ряду.

Послідовність коефіцієнтів автокореляції першого, другого і т.д. порядків утворює автокореляційну функцію часового ряду. Коефіцієнт автокореляції *j*–го порядку визначає ступінь залежності між спостереженнями, які знаходяться на відстані j періодів. Графік цієї залежності (*корелограма*) представляє деяку криву, що показує як змінюється взаємовплив між спостереженнями в залежності від відстані між ними в часі. Тлумачення значень функції автокореляції та її поведінки не є легкою справою. Проте, аналіз автокореляційної функції та її графіка (*корелограми*) дозволяє виявити особливості динаміки системи:

- якщо найбільш високим виявляється *додатній* коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тренд (тенденцію до зростання чи спадання);
- якщо найбільш високим виявляється *від'ємний* коефіцієнт автокореляції першого порядку, досліджуваний ряд є реверсивним (схильним до коливань);
- якщо найбільш високим виявляється *додатній* коефіцієнт автокореляції порядку T, ряд містить циклічні коливання з періодичністю T;
- якщо жоден з коефіцієнтів автокореляції не є

статистично значущим, можна зробити припущення, що даний ряд не містить тенденції та циклічних коливань і має структуру схожу до структури випадкового процесу.

У практичних дослідженнях значущим вважають коефіцієнт автокореляції, модуль якого перевищує значення $0{,}2$.

Для побудови корелограми в середовищі MS Excel можна використовувати функцію **КОРРЕЛ()**.

Наприклад, функція «КОРРЕЛ(b4 : b50 ; b5 : b51)» дозволяє отримати значення першого коефіцієнта автокореляції ρ_1 для ряду довжиною 56 елементів, розміщеного в комірках b4 : b59; функція «КОРРЕЛ(b4 : b50 ; b6 : b52)» – значення другого коефіцієнта автокореляції ρ_2 тощо.

Для отримання більш точної корелограми можна використати один із комп’ютерних статистичних пакетів, наприклад *Statistica* (рис. 10.1). Аналіз корелограми, представленої на рис. 10.1, дозволяє ідентифікувати такі властивості часового ряду врожайності, як реверсивність та слабко виражена циклічність.

Національний університет
водного господарства
та природокористування

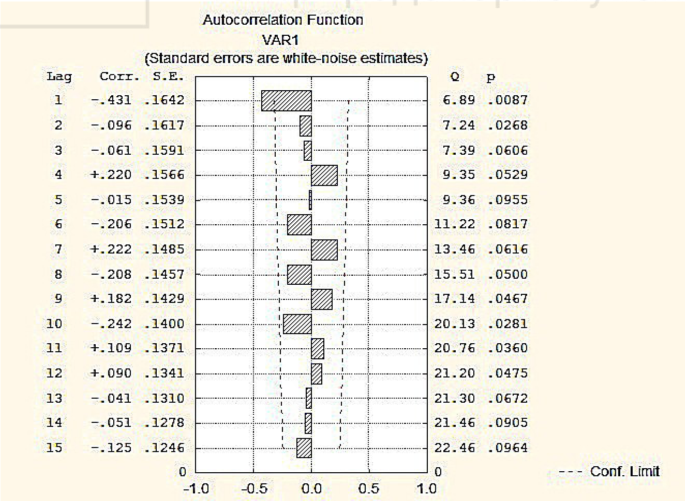


Рис. 10.1. Корелограма ряду різниць врожайності зернових культур

10.2. Дослідження систем методами хаотичної динаміки

Найпростішим типом динамічної поведінки системи є лінійна динаміка, яка характеризується рівномірним зростанням чи спаданням характеристичного показника. Для дослідження систем з лінійним типом динаміки доцільно використовувати трендові моделі та регресійні моделі. Більш складною є періодична динаміка, для якої є характерними регулярні періоди зростання та спадання показника. Для дослідження таких систем використовують метод гармонічних функцій, або ж систему диференціальних рівнянь (наприклад, модель «хижак–жертва»). Найбільш складним типом динаміки є ***хаотична динаміка***, при якій тривалість та амплітуда коливань є змінними (*наприклад*, система Лоренца).

Для вивчення хаотичної динаміки розроблено низку методів, які можуть бути корисними, зокрема, і для ідентифікації систем за даними часових рядів. Оскільки більшість часових рядів є нестаціонарними, спочатку необхідно змоделювати та вилучити з ряду його тренд та циклічну компоненту. Отриманий ряд залишків необхідно згладити, щоб відсіяти сторонні шуми. Дослідження згладженого ряду залишків дозволяє побудувати фазовий портрет системи, обчислити розмірність системи, оцінити ступінь її прогнозованості.

Побудова фазового портрету. Використання методів хаотичної динаміки вимагає великої довжини досліджуваного ряду (близько 1000 елементів). У випадку, якщо досліджувані ряди є коротшими, використовують ергодичну гіпотезу і моделюють один довгий ряд низкою більш коротких рядів, для яких виконується властивість однорідності.

Наприклад, при дослідженні динаміки врожайності зернових в Україні довжина часових рядів складає 57 елементів. Однак, наявність циклів приблизно однакової тривалості дозволяє розглядати кожен ряд як декілька витків фазової траєкторії. Такі ряди містять три цикли (три витки фазової траєкторії) середньою тривалістю 19 років. Отже можна вважати ці ряди однорідними. Розглянувши групу 15 областей з однорідною динамікою врожайності, отримаємо 45 витків фазової траєкторії тривалістю 19 років. Згідно з ергодичним принципом такий набір часових рядів можна розглядати як аналог одного часового ряду протяжністю 855 елементів. Отриманий ряд можна досліджувати методами нелінійної динаміки.

Дослідження складної нелінійної системи починається з побудови фазового портрета системи, яка характеризується часовим рядом одного показника $\{x_1, x_2, \ldots, x_n\}$. Але, для побудови фазового портрета необхідно мати мінімум дві змінні. Якщо розглянути два відрізки часового ряду, зміщені один відносно одного на деякий лаг τ, то їх умовно можна розглядати як різні змінні. Це дозволяє отримати двовимірну проекцію фазового портрету досліджуваної системи. Декілька відрізків часового ряду, зміщених на лаг τ, дають змогу отримати *n*-вимірну проекцію фазового портрету. Такий метод побудови фазового портрету називають **реконструкцією** фазового портрету за даними часового ряду, або ж зануренням часового ряду у фазовий простір.

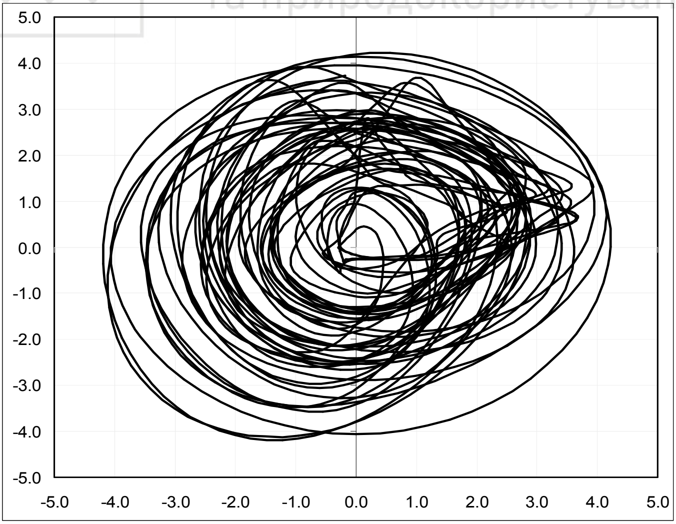


Рис. 10.2. Двовимірна проекція фазового портрету системи зерновиробництва України

Вигляд фазового портрету дозволяє оцінити характер динаміки системи. Якщо фазова траєкторія є замкнутою, це свідчить про циклічну поведінку системи. Якщо фазова траєкторія наближається до стаціонарної точки, це свідчить про стійку динаміку системи. Якщо ж фазові траєкторії віддаляються від стаціонарної точки, це є свідченням нестійкої динаміки системи.

На рис. 10.2 зображено двовимірну проекцію фазового портрету системи зерновиробництва України, отриманого із сукупності згладжених рядів залишків рядів врожайності. Поведінка фазової траєкторії системи демонструє яскраво виражену циклічність. На перший погляд може здатися, що в деяких точках траєкторія перетинає саму себе. Однак, відомо, що для динамічної системи це неможливо. Причина полягає у тому, що ми розглядаємо двовимірну проекцію фазового портрету. При більшій розмірності простору фазова траєкторія не буде самоперетинатися. Щоб це «побачити», розмірність фазового простору повинна бути достатньо великою.

Визначення розмірності системи. Наступним кроком аналізу системи є визначення її розмірності, тобто визначення мінімальної кількості параметрів, яка необхідна для описання динаміки системи. Для цього необхідно побудувати фазовий простір такої розмірності, у якому фазові траєкторії системи ніколи не самоперетинаються. Якщо розмірність такого фазового простору є скінченною (на практиці вважають, що вона не повинна перевищувати 5 – 6) – це вказує на детермінований характер поведінки системи. Велике (нескінченне) значення розмірності є свідченням випадкового характеру поведінки системи.

Мінімальну розмірність фазового простору, при якій фазовий портрет системи вдається відобразити адекватно, називають **розмірністю вкладення**. Розмірність вкладення є оцінкою розмірності динамічної системи за даними часового ряду.

Для знаходження розмірності вкладення використовують метод «фальшивих сусідів». При цьому використовують поняття *близькості* фазових векторів. Для визначення відстані між векторами $(x_{i1}, x_{i2}, \ldots, x_{iD})$ та $(x_{j1}, x_{j2}, \ldots, x_{iD})$, зазвичай, використовують евклідове визначення відстані

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{iD} - x_{jD})^2} . \qquad (10.4)$$

Якщо відстань між двома точками фазового простору є малою, такі точки вважають близькими («сусідами»). При збільшенні розмірності фазового простору D в результаті появи нових ненульових координат відбувається збільшення відстані між двома близькими точками простору – ефект «розтягування фазового простору». Якщо при збільшенні розмірності фазового простору на 1 відстань між двома близькими точками збільшилась у декілька разів, такі точки називають «фальшивими сусідами». Велика кількість фальшивих сусідів свідчить про те, що розмірність фазового простору є занадто малою для адекватного відображення фазового портрета динамічної системи і її необхідно збільшити.

Процедура оцінювання розмірності системи починається з розмірності 2. Вибирається група векторів фазового простору, відстань між якими є найменшою – «найближчі сусіди». Потім розмірність збільшується на 1 і розраховується коефіцієнт k збільшення відстані між векторами – «найближчими сусідами». За розмірність системи вибирають таке число D, за якого при переході до розмірності $D+1$ відстань між «найближчими сусідами» збільшувалась незначним чином (менше 10%).

Наприклад, на рис. 10.3 видно, що при $D > 4$ зростання відстані між «векторами–сусідами» є незначним. Тому розмірність вкладення системи зерновиробництва України можна оцінити значенням 4.

Показники Ляпунова. Фундаментальною характеристикою динамічної системи є набір *показників Ляпунова*, які описують поведінку траєкторій у фазовому просторі. В математичному сенсі показники Ляпунова – це власні значення системи рівнянь (матриці коефіцієнтів системи), що моделює відповідну економічну, природничу чи інформаційну систему. Кожна система розмірності D має D показників Ляпунова $\lambda_1, \lambda_2, ..., \lambda_D$. Ці показники можуть мати від'ємні, додатні чи нульові значення.

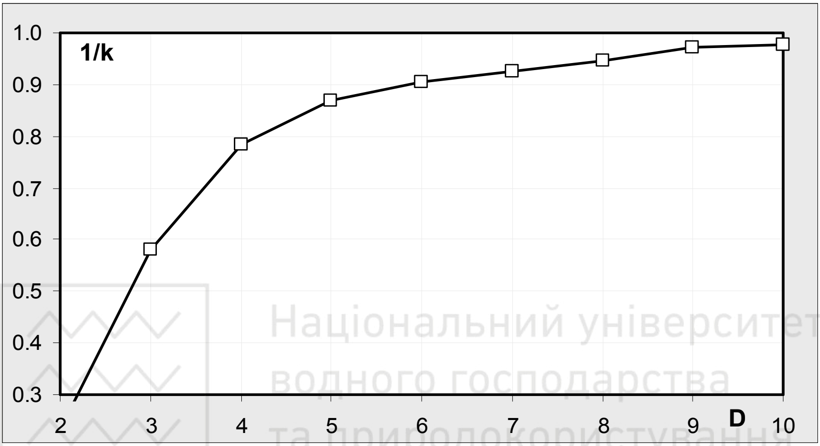


Рис. 10.3. Розмірність вкладення D для фазового портрету системи зерновиробництва України

Кожний показник відображає середню швидкість розбігання ($\lambda > 0$) або сходження ($\lambda < 0$) початково близьких фазових траєкторій системи. Найбільш важливим є перший (найбільший) показник Ляпунова, що відповідає максимальному власному значенню системи.

Значення максимального показника Ляпунова дозволяє ідентифікувати тип динаміки системи:

- при $\lambda_{max} < 0$ система еволюціонує до стану рівноваги,
- при $\lambda_{max} = 0$ поведінка системи є циклічною,
- у випадку, коли $\lambda_{max} > 0$, система характеризується хаотичною динамікою.

Знання максимального показника Ляпунова системи є дуже важливим, оскільки він визначає ***можливість прогнозування*** розвитку системи. У перших двох випадках система є прогнозованою, причому ми можемо прогнозувати її розвиток на довільно великий проміжок часу. За наявності хаотичної динаміки в системі максимальний час прогнозування обмежений внаслідок швидкого розбігання фазових траєкторій. Значення максимального прогнозного горизонту T_{max} визначається із співвідношення

$$T_{max} = 1/\lambda_{max}.$$

Наприклад, для системи зерновиробництва України (рис. 10.2) значення максимального показника Ляпунова становить $\lambda_{max} = 0{,}05$ і, отже, максимальний горизонт прогнозування – $T_{max} = 20$ років.

10.3. Регресійні та авторегресійні моделі динаміки систем

Вибір методу прогнозування багато в чому залежить від обсягу та структурного складу вихідних даних. Якщо є доступними дані про декілька параметрів системи за певний часовий період, застосовують кореляційний аналіз вхідних даних. Якщо встановлена кореляційна залежність між досліджуваною величиною y_t і впливаючими параметрами $x_1, x_2, \ldots, x_p$, це служить підставою для побудови регресійних прогнозних моделей виду

$$y_t = a_0 + a_1 x_{1t} + \ldots + a_p x_{pt}. \qquad (10.5)$$

При побудові моделі (10.5) необхідно визначити відсутність внутрішньої кореляції між змінними $x_1, x_2, \ldots, x_p$. Якщо присутня кореляція між параметрами x_i та x_k, одну з цих змінних необхідно вилучити з моделі.

В системах часто зустрічаються випадки, коли на сьогоднішній стан процесу впливають вчорашні значення чинника. Побудова звичайних лінійних регресій не може адекватно змоделювати та дати повне уявлення про напрями розвитку таких часових рядів. У цих випадках використовують моделі авторегресії, у яких роль пояснюючих чинників виконують попередні спостереження часового ряду. Модель часового ряду, у якій значення поточного спостереження залежать від попередніх значень спостережень за p періодів, називається авторегресійною моделлю, або ***АР–моделлю*** порядку p.

Авторегресійна модель порядку p має вигляд

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots\ldots + a_p y_{t-p}, \qquad (10.6)$$

або з використанням лагового оператора L

$$y_t = a_0 + a_1 L y_t + a_2 L^2 y_t + \ldots\ldots + a_p L^p y_t.$$

Окремі коефіцієнти АР–моделі можуть бути рівними нулю. Для визначення оптимальної структури авторегресійної моделі необхідно побудувати авторегресійну функцію ряду перших різниць. Якщо, наприклад, авторегресійна функція має значущі коефіцієнти першого і четвертого порядку (рис. 10.1) авторегресійна модель матиме вигляд

$$y_t = a_0 + a_1 y_{t-1} + a_4 y_{t-4}. \qquad (10.7)$$

10.4. Прогнозування методом «найближчих сусідів»

Ефективним підходом до прогнозування динаміки складних систем є модель найближчих сусідів. Спочатку будується фазовий простір системи необхідної розмірності D. Фазова траєкторія складається з точок фазового простору, кожна з яких має D координат. Основна ідея моделі найближчих сусідів полягає в тому, що близькі фазові вектори на короткому відрізку часу еволюціонують однаково.

Задача прогнозу формулюється наступним чином. Дано послідовність фазових векторів $\{y_i\}$ $\left(i = \overline{1,n}\right)$; необхідно змоделювати вектор y_{n+1}. Для того, щоб оцінити зміну фазового вектора y_n, необхідно знайти m найближчих до нього векторів (найближчих сусідів). Позначимо ці вектори $y_{n1}, y_{n2}, ..., y_{nm}$. В процесі еволюції системи ці вектори переходять у наступні за ними вектори – послідовності наступних елементів часового ряду $y_{n1+1}, y_{n2+1}, ..., y_{nm+1}$.

Як найпростішу модель прогнозного вектора y_{n+1} можна використати вектор

$$y_{n+1} = \frac{1}{m}(y_{n1+1} + y_{n2+1} + ... + y_{nm+1}), \qquad (10.8)$$

тобто прогноз отримується усередненням прогнозів k–найближчих векторів–сусідів.

Більш точна модель використовує зважене усереднення окремих прогнозів

$$y_{n+1} = \frac{w_1 y_{n1+1} + w_2 y_{n2+1} + ... + w_m y_{nm+1}}{w_1 + w_2 + .. + w_m}. \qquad (10.9)$$

В ролі ваг w_i здебільшого використовують обернену відстань векторів-сусідів до прогнозованого вектора y_n. Ефективність застосування моделі найближчих сусідів залежить від вдалого вибору параметрів m і D. Кількість сусідів m не може бути занадто малою, бо це спричиняє великий розкид значень прогнозів, і занадто великою, бо це може привести до зміщення від бази прогнозу.

Розглянемо процес прогнозування ряду з використанням методу найближчих сусідів на *прикладі*.

Нехай деяка система характеризується часовим рядом одного з параметрів $\{x_1, x_2, \ldots, x_{54}\}$. Вважатимемо, що розмірність системи $D = 4$. Отже, стан системи в даний момент описується набором координат (4-вимірним вектором): $(x_{51}, x_{52}, x_{53}, x_{54})$. Необхідно побудувати прогноз на наступний момент часу, тобто спрогнозувати значення параметра x_{55}. За методом найближчих сусідів визначаємо декілька наборів (векторів), які є найближчими до вектора поточного стану системи $(x_{51}, x_{52}, x_{53}, x_{54})$. Нехай перший такий набір (вектор) має вигляд (x_1, x_2, x_3, x_4). Йому відповідає частковий прогноз x_5. Беремо інший «близький» вхідний набір, наприклад, $(x_{11}, x_{12}, x_{13}, x_{14})$. Йому відповідає частковий прогноз x_{15}. Продовжуючи таким чином визначаємо всі m часткових прогнозів і за співвідношенням (10.8) або (10.9) будуємо остаточний прогноз. Коли стане відомим фактичне значення параметра системи x_{55}, тоді можна визначити похибку прогнозу.

За допомогою вищеописаної методики можна отримувати прогнозні значення, взагалі кажучи, на будь-який період. Але будь-який прогноз рідко може бути ідеально точним. Частіше за все дослідники вказують деякий надійний інтервал, куди повинно попасти реальне значення у відповідний період. За допомогою часових рядів, як уже зазначалося, можна будувати інтервальні прогнози з різним рівнем надійності. Ідея побудови інтервалів базується на тому, що є деяке відхилення на однакову визначену величину у напрямі зростання та спадання від точного прогнозного значення. Основним припущенням для реалізації цієї ідеї є нормальність збурень, за якої з певним рівнем надійності можна розрахувати ймовірність попадання прогнозного значення в симетричний довірчий інтервал. Таким чином, знаючи напевно збурення минулих періодів, можна сформувати прогноз та надійний інтервал на майбутні періоди.

Проблема в тому, що, як правило, ці збурення нам невідомі. Це вимагає застосування адекватного методу оцінки збурень в попередні періоди.

10.5. Прогнозування методом штучних нейронних мереж

Штучні нейронні мережі застосовують у таких областях як розпізнавання та класифікація образів, моделювання та прогнозування складних процесів, обробка сигналів.

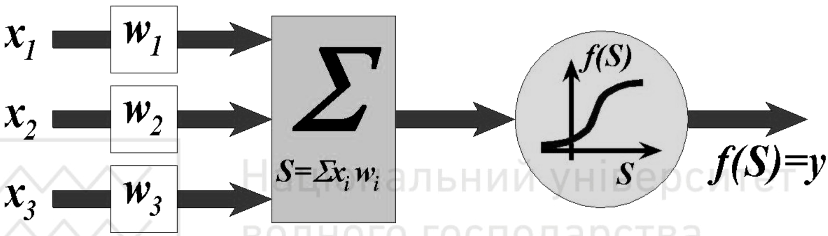


Рис. 10.4. Схема штучного нейрона з трьома входами

Нейроподібний елемент, який використовується у моделюванні нейронних мереж, зображено на рис. 10.4. На вхід нейроподібного елемента поступає набір вхідних сигналів $x_1, x_2, ..., x_n$, який є набором вихідних сигналів інших нейроподібних елементів.

Кожний вхідний сигнал $x_1, x_2, ..., x_n$ має неоднакову важливість для поведінки системи. Тому, його потрібно помножити на вагу відповідного зв’язку $w_1, w_2, ..., w_n$. Зважені вхідні сигнали поступають на блок сумації, який здійснює алгебраїчну сумацію і визначає рівень збудження нейроподібного елемента V

$$V = \sum_{i=1}^{n} w_i x_i \; . \qquad (10.10)$$

Вихідний сигнал нейрона y визначається шляхом пропускання рівня збудження V через деяку нелінійну активаційну функцію $f(V)$: $y = f(V)$.

Головною перевагою моделі нейромереж перед іншими методами прогнозування є їх здатність до ***навчання***. Для навчання нейронної мережі необхідні навчаючі дані. Навчаючі дані – це множина еталонних пар «вхід–вихід». Після завершення періоду навчання нейронна мережа може бути використана для прогнозування, оскільки «навчена» відтворювати вихідні сигнали, які динамічно відповідають відповідним вхідним сигналам.

Прогнозування часового ряду за методом нейронних мереж зводиться до апроксимації функції багатьох змінних за заданим набором прикладів. Нейромережу використовують для відновлення цієї функції за набором векторів, побудованих на основі даних часового ряду. Значення D визначає ширину вікна, яке ковзає вздовж часового ряду, задаючи вхідні набори. Оптимальна ширина вхідного вікна D співпадає з розмірністю системи, визначеної методами хаотичної динаміки.

Процедура прогнозування має наступний вигляд. Нехай, *наприклад*, система характеризується часовим рядом одного з параметрів $\{x_1, x_2, \ldots, x_{54}\}$. Розмірність системи $D = 4$. Отже, стан системи в даний момент описується набором координат $(x_{51}, x_{52}, x_{53}, x_{54})$. Необхідно побудувати прогноз на наступний момент, тобто спрогнозувати значення параметра x_{55}. Визначаємо початкові значення ваг $\{w_1, w_2, w_3, w_4\}$ випадковим чином, зберігаючи умову $w_i < 1$. Вводимо у нейронну мережу перший стан системи (x_1, x_2, x_3, x_4). Отримуємо частковий прогноз p_1. Враховуючи фактичне значення параметра x_5, визначаємо похибку $e_1 = p_1 - x_5$. Формуємо наступний вектор (x_2, x_3, x_4, x_5), вводимо його у нейронну мережу і отримуємо прогноз p_2, якому відповідає похибка e_2. Продовжуємо таким чином, поки не досягнемо передостаннього вектора $(x_{50}, x_{51}, x_{52}, x_{53})$. Визначаємо загальну похибку прогнозу $E = \sum_i e_i^2$.

Перенормовуємо початкові ваги так, щоб зменшити похибку прогнозу. Для цього використовуємо один з методів мінімізації функцій багатьох змінних, наприклад, метод градієнтного спуску. Після цього процедура прогнозування починається спочатку. Загальна похибка прогнозу на другому кроці є меншою від початкової похибки. Продовжуємо таким чином до того часу, поки загальна похибка не стане меншою від наперед заданої величини (наприклад, $\varepsilon = 0{,}01$). На цьому процедура навчання мережі закінчується і можна приступати до прогнозування. Описаний метод носить назву «метод зворотного поширення похибки».

Теми для роздумів і контрольні запитання

1. У чому полягає складність опису нелінійних динамічних процесів?
2. Поняття про детермінований хаос. Хаотична динаміка.
3. Кусково-лінійні та порогові моделі для опису нелінійних динамічних процесів. Переваги і недоліки.
4. Що таке ідентифікація системи?
5. Визначення ступеня випадковості системи за критерієм Кендала.
6. Який часовий ряд називають персистентним, антиперсистентним, випадковим?
7. Які типи моделей доцільно використовувати для прогнозування персистентних часових рядів?
8. Які типи моделей доцільно використовувати для прогнозування антиперсистентних часових рядів?
9. Які типи моделей можна використовувати для прогнозування стохастичних часових рядів?

10. В чому суть кореляційного аналізу часових рядів?
11. Автокореляція рівнів часового ряду, коефіцієнт автокореляції, лаг між рівнями ряду.
12. Що показує корелограма часового ряду?
13. В чому полягає метод реконструкції фазового портрету за даними часового ряду?
14. Що таке розмірність динамічної системи і як вона визначається?
15. Знаходження розмірності вкладення динамічної системи за методом «фальшивих сусідів».
16. Показники Ляпунова і їх використання для прогнозування динаміки системи.
17. Поняття про авторегресійну модель.
18. Прогнозування методом «найближчих сусідів»
19. Поняття про штучні нейронні мережі.
20. В чому полягає суть навчання нейронної мережі?

11. МЕТОДИ ПРИЙНЯТТЯ РІШЕНЬ

11.1. Проблема прийняття рішень і узгодження цілей

У процесі управління особі, яка приймає рішення, (наприклад, керівнику підприємства чи організації) доводиться приймати рішення в умовах багатоваріантності розвитку системи (фірми, підприємства). ***Управління системою*** – це послідовність прийняття рішень та принципи, за якими вони приймаються. Головною метою будь-якого управління системою є підвищення ефективності її функціонування. Для забезпечення ефективності рішень необхідно мати експертне, логічне, а найкраще, формально-математичне обґрунтування ефективності рішень.

Методи прийняття рішень використовуються у різних галузях управління: проектуванні складних технічних і організаційних систем, плануванні розвитку міст, доборі програм розвитку економіки і енергетики регіонів, організації нових економічних зон тощо. Потреба в застосуванні засобів і методів прийняття рішень в управлінні очевидна: швидкий розвиток і ускладнення економічних зв’язків, виявлення залежностей між окремими складними процесами та явищами, які раніше здавалися не пов’язаними один з одним, проте, призводять до різкого зростання труднощів під час прийняття обґрунтованих рішень. Витрати на прийняття рішень зростають, наслідки помилок стають усе серйознішими, а звернення до фахового досвіду та інтуїції не завжди зумовлює вибір найкращої стратегії. Застосування методів прийняття рішень дає змогу розв’язати цю проблему, до того ж швидко та достатньо точно і ефективно.

Однією з умов ефективності прийняття рішень є наявність достатньої інформації. «Хто володіє інформацією – той володіє світом» (Уїнстон Черчіль). Часто в процесі економічної діяльності для здобуття необхідної інформації вдаються до не зовсім чесних прийомів (промислове шпигунство). Тому важливим аспектом діяльності підприємства чи організації є захист інформації і забезпечення недоступності особливо важливої інформації для сторонніх осіб Загалом, рівень інформаційного забезпечення прийняття рішення залежить від низки чинників, зокрема, це якість інформації про проблемну ситуацію, наявні ресурси щодо збору, обробки і зберігання даних, прикладені зусилля, час для збору інформації тощо.

Іншою умовою ефективності управління є співвідношення ступенів складності керуючої системи і керованої системи. Коли керівник приймає рішення, перед ним є множина варіантів прийняття рішень. Для вибору оптимального варіанту керівник повинен застосувати множину методів аналізу і оцінки цих варіантів. Уільям Ешбі довів теорему, яка відома як **принцип необхідної різноманітності**. *Для успішного вирішення задачі управління управляюча система повинна мати більшу різноманітність, ніж об'єкт управління.*

Проблема прийняття рішення передбачає наявність декількох можливих варіантів дій (альтернатив). Прийняття рішення – це дія над множиною альтернатив, у результаті якої вибирається оптимальна в деякому відношенні альтернатива.

Процедура прийняття рішення включає такі етапи:

- описання проблеми;
- постановка задачі;
- аналіз наявної інформації;
- розробку альтернатив;
- вибір оптимальної альтернативи.

У практичних задачах прийняття рішень формалізація кожного з вищенаведених кроків процесу прийняття рішень пов'язана з певними, іноді дуже складними, проблемами. В першу чергу, постає проблема визначення мети та засобів її досягнення. Адже, можна поставити апріорі недосяжні або навіть абсурдні чи злочинні цілі («моя мета – пробігти стометрівку за п'ять секунд», «наша мета – комунізм», «наша мета – чистота раси» тощо). Можна використовувати нецивілізовані, навіть злочинні, методи досягнення цілком досяжної мети (мета – «стати президентом», «стати багатим», «отримати п'ятірку на іспиті» тощо).

Припустимо, що мета і методи її досягнення визначені. Постає проблема побудови множини альтернатив – варіантів дій, направлених на досягнення мети. Тут, у першу чергу, виникає проблема побудови «повного списку» альтернатив. Можлива ситуація, коли не включення певної альтернативи призведе до неможливості розв'язання задачі або до її «неякісного» розв'язання. Так, не включення до економічної системи колишнього СРСР ринкових механізмів призвело до деградації суспільства і розпаду СРСР.

Не менш складною є проблема оцінки альтернатив – «до чого приведе та чи інша вибрана дія». Як правило, оцінки альтернатив мають суб'єктивний характер, бо отримуються на основі обробки експертної інформації. Навіть якщо можна оцінювати альтернативи за допомогою «об'єктивних» процедур (наприклад, вимірювати вагу чи вартість товару, відстань між населеними пунктами), постає проблема визначення всіх або хоча б «найважливіших» аспектів оцінки кожної альтернативи. Тут також неврахування навіть одного аспекту в оцінці варіантів дії може призвести до катастрофічних наслідків (згадаймо Чорнобильську катастрофу – неврахування ризику аварії при будівництві АЕС привело до трагічних результатів).

І ще одна принципова складність – вибір принципу порівняння альтернатив і на його основі – принципу оптимальності. Якщо на попередньому етапі визначені кількісні оцінки альтернатив, то вибір принципу оптимальності зводиться до вибору критерію оптимізації, який максимально відповідає меті. Наприклад, якщо для тренера футбольної команди мета – перемога у наступному матчі, то за принцип оптимальності може слугувати такий критерій: «Перемагає та команда, яка виконує за матч більшу сумарну кількість успішних тактично-технічних елементів (передач м'яча, відборів, ударів по воротах тощо)» (В. Лобановський). Якщо мета задачі прийняття рішень описується декількома кількісними критеріями (і, отже, маємо задачу багатокритеріальної оптимізації), необхідно визначити – на основі якого «глобального» принципу оптимальності будуть порівнюватись (і вибиратися кращі) альтернативи [5, С. 6–7].

У процесі розв'язання загальної задачі прийняття рішень, як правило, беруть участь три групи осіб: особи, що приймають рішення (ОПР), експерти та консультанти.

Особами, що приймають рішення, називають керівників (або колективні органи, наприклад, Вчену раду університету), що формулюють постановку задачі та пошук її розв'язання і несуть відповідальність (моральну, юридичну, фінансову тощо) за наслідки прийнятих рішень; визначають, які засоби є допустимими (недопустимими) для досягнення мети.

Експерт – це спеціаліст у своїй галузі, що володіє інформацією про задачу, але не несе прямої відповідальності за результати її розв'язання. Експерти допомагають ОПР на всіх стадіях постановки й розв'язання проблеми.

Аналітиками (консультантами, дослідниками) називають спеціалістів із теорії прийняття рішень. Вони розробляють модель (математичну, інформаційну чи іншу) задачі прийняття рішення, процедури прийняття рішення, організовують роботу ОПР і експертів.

У найпростіших ситуаціях ОПР може виступати одним у трьох ролях, у більш складних – ОПР може поєднувати функції аналітика, звертаючись до спеціалістів із вузьким профілем (експертів) для вирішення часткових проблем. У загальному випадку ОПР залучає до вирішення проблем аналітиків-консультантів, які, у свою чергу, звертаються до експертів.

До основних чинників, що впливають на прийняття управлінських рішень, можна віднести наступні:

- чинники, що утворюють структуру проблеми, яка вимагає рішення (цілі самого проекту, особисті цілі керівника, види альтернативних рішень і їх наслідки, обсяги та вартість різного роду ресурсів, темпи поширення інформації, час, відведений на прийняття рішення, тип реакції ОПР тощо);
- чинники, що перебувають поза проблемою (управлінські процедури і правила, політика і стратегія організації, юридичні рамки діяльності, тип і характеристики зовнішнього середовища).

Нагромаджений практичний досвід у галузі проблем прийняття рішень показує, що часто найважчими і найважливішими виявляються такі аспекти цих проблем, які безпосередньо не стосуються процесу прийняття рішення, а саме:

• формулювання проблеми і введення ОПР та експертів в проблематику задач, які потрібно розв'язати;

• формування спільної мови спілкування для різних груп експертів і ОПР, узгодження їхніх думок і поглядів;

• виявлення справжніх цілей розв'язання та постановки задачі.

Проблема узгодження цілей

Складність процесу управління системою, в першу чергу, пов'язана з проблемою: як вибрати найбільш ефективний спосіб управління системою? На перший погляд відповідь очевидна: потрібно надати якнайширшу свободу дій для всіх елементів системи. Проте, в такому випадку можуть виникати протиріччя між цілями окремих елементів системи і їх неузгодженість з генеральною метою системи.

Приклад. Розглянемо діяльність деякої фірми, що виробляє певні види продукції і прагне отримати максимальний прибуток від її продажу. Питання, що потребує вирішення: скільки готової продукції слід зберігати на складі підприємства і скільки різновидів її слід виробляти? Розглянемо «особисті» інтереси різних відділів фірми і відразу ж виявимо їх розбіжність.

• Виробничий відділ буде зацікавлений у тривалому і безперервному виробництві одного і того ж виду продукції. В цьому випадку будуть найменшими витрати на налаштування устаткування.

• Відділ збуту, навпаки, буде відстоювати ідею виробництва максимального асортименту продукції і великих запасів на складах. Це сприятиме збільшенню доходів від продажу.

• Фінансовий відділ буде наполягати на мінімумі складських запасів, оскільки їх зберігання вимагає затрат і не приносить прибутку.

• Відділ кадрів вважатиме за потрібне виробляти продукцію стабільно (навіть в періоди ділового спаду) і в одному, і тому ж асортименті, так як в цьому випадку не буде проблем плинності кадрів.

Отже, для управління такою системою необхідно здійснювати узгодження локальних цілей окремих підрозділів з досягненням глобальної мети підприємства – отримання максимального прибутку.

Проблему узгодження цілей окремих підсистем системи можна, в загальному, змоделювати у вигляді

$$E = f(X, Y), \tag{11.1}$$

де E – деякий кількісний показник ефективності системи (критерій ефективності); X – керовані змінні системи – ті, на які ми можемо впливати; Y – некеровані змінні (стани середовища).

Якщо стан середовища є несуттєвим – кажуть про прийняття управлінських рішень в умовах визначеності. Якщо ж із впливами навколишнього середовища ми змушені рахуватися, то рішення приймаються в умовах невизначеності.

11.2. Класичні задачі прийняття управлінських рішень

Задачі розподілу ресурсів. У цих задачах доводиться виконувати декілька операцій з продукцією і, при цьому, не вистачає ресурсів або обладнання для виконання всіх цих операцій. Необхідно знайти спосіб найбільш ефективного виконання операцій з урахуванням обмежень на ресурси.

Існує цілий клас задач системного аналізу, де мова йде про необхідність мінімізувати деяку функцію такого вигляду

$$E = a_1x_1 + a_2x_2 + ... + a_nx_n, \tag{11.2}$$

де x_i – шукані змінні, a_i – відповідні їм коефіцієнти або «ваги змінних», і при цьому мають місце обмеження на змінні (обмеження ресурсів).

Об’єднують всі такі задачі методи їх вирішення – методи математичного програмування.

Задачі математичного програмування. Математичне програмування – один з основних інструментів управління системами, що полягає в вивченні методів розв’язування оптимізаційних задач та дослідження отриманих розв’язків.

Як самостійний науковий напрямок математичне програмування сформувалось на початку 40-х років XX ст. В 1939 р. математик ***Л. В. Канторович*** опублікував роботу «Математичні методи організації та планування виробництва», в якій сформулював принципово новий клас екстремальних задач з обмеженнями і розробив ефективний метод їх розв’язання. Так було започатковано новий розділ прикладної математики, який отримав назву «лінійне програмування». Дослідження Канторовича в цій галузі сприяли створенню строго наукового інструментарію для розв’язання фундаментальних економічних проблем (ефективності капіталовкладень, ціноутворення, теорії ренти тощо), за що він (разом з Т. Ч. Купмансом) у 1975 р. був удостоєний Нобелевської премії з економіки.

Американський економіст ***Т. Купманс*** протягом багатьох років привертав увагу математиків до низки задач, пов’язаних з військовою тематикою, для вирішення проблем знаходження екстремумів лінійних функцій на багатогранниках, що задаються лінійними нерівностями. Саме за пропозицією Купманса цей розділ математики одержав назву ***лінійного програмування***.

Термін «програмування» («математичне програмування», («динамічне програмування») тут не має нічого спільного з комп’ютерами чи програмуванням у сучасному розумінні, а виник почасти через історичне непорозуміння, почасти через неточний переклад з англійської мови: ***mathematical programming***, що означає розроблення на основі математичних розрахунків програми дій для досягнення обраної мети. З програмуванням для ЕОМ математичне програмування має лише те спільне, що більшість виникаючих на практиці задач математичного програмування занадто громіздкі для ручного розрахунку, тому розв’язати їх можна тільки за допомогою комп’ютера, попередньо склавши програму.

У 1947 р. ***Дж. Данцигом*** був розроблений основний метод розв'язування задач лінійного програмування – ***симплексний метод***. Наступним кроком стали праці Дж. Неймана (1947 р.) щодо розвитку концепції двоїстості, що уможливило розширення практичної сфери застосування методів лінійного програмування.

У п'ятдесяті роки XX ст. з'явилися розробки нових алгоритмів, теоретичні дослідження з різних напрямків математичного програмування: 1951 р. – праця Г. Куна і А. Таккера, в якій наведено необхідні та достатні умови оптимальності нелінійних задач; 1954 р. – Чарнес і Лемке розглянули наближений метод розв'язання задач з сепарабельним опуклим функціоналом та лінійними обмеженнями; 1955 р. – низка робіт, присвячених квадратичному програмуванню; сформувався новий напрямок математичного програмування – динамічне програмування, значний внесок у розвиток якого вніс американський математик Р. Белман.

В загальному випадку ***задача математичного програмування*** полягає у відшуканні екстремуму (мінімуму або максимуму) деякої функції (цільової функції або функції мети)

$$f(x_1, x_2, \ldots, x_n) \to \max(\min), \tag{11.3}$$

за обмежень

$$g_i(x_1, x_2, \ldots, x_n) \le b_i, \;\; i = \overline{1, m}, \tag{11.4}$$

де f, g_i – задані функції; b_i – деякі задані числа.

В залежності від властивостей функцій f, g_i математичне програмування можна класифікувати на:

- лінійне програмування (якщо f, g_i – лінійні функції);
- нелінійне програмування (якщо f або g_i – нелінійні функції);

- цілочисельне програмування (якщо шукані x_i мають бути цілими числами);
- параметричне програмування (якщо f або g_i містять випадкові величини);
- динамічне програмування (процес розв’язування задачі багатоетапний).

Кожна система має мету розвитку та функціонування. Це може бути, наприклад, отримання максимального прибутку або мінімізація витрат. Вважатимемо, що ступінь досягнення мети має кількісну міру. Тоді математичну модель системи можна побудувати наступним чином.

1. Опишемо величини, які будуть використовуватись в моделі:

- *параметри моделі* c_k $(k = 1, 2, \ldots, l)$ – числові характеристики системи, які є сталими величинами;
- *керовані змінні* x_i $(i = 1, 2, \ldots, n)$ – незалежні змінні характеристики моделі, які можуть змінюватись в певному інтервалі;
- *некеровані змінні* y_j $(j = 1, 2, \ldots, m)$ – незалежні змінні, значення яких носять об’єктивний характер (найчастіше їх значення залежать від зовнішнього середовища);
- величина z, яка вимірює ступінь досягнення мети системи.

2. Встановимо взаємозв’язки між величинами:

- *цільова функція* $z = f(x_1, \ldots, x_n; y_1, \ldots, y_m; c_1, \ldots, c_l)$ – функціональна залежність, яка встановлює взаємозв’язок між керованими та некерованими змінними і параметрами моделі;

– *система обмежень*

$$\begin{cases} \varphi_1(x_i, y_j, c_k) \le (\ge, =, <, >) 0, \\ \ldots \\ \varphi_s(x_i, y_j, c_k) \le (\ge, =, <, >) 0, \end{cases}$$

виражає в математичній формі обмеження на вибір значень керованих змінних x_i, що обумовлені зовнішніми щодо досліджуваної системи умовами (параметрами системи).

Задача математичного програмування полягає у знаходженні таких значень керованих змінних x_i, при яких цільова функція z набуває екстремального (максимального чи мінімального) значення.

Будь який набір значень змінних $(x_1, \ldots, x_n)$, що задовольняє систему обмежень, називається *допустимим* (або *опорним*) *розв'язком* (*планом*). Сукупність усіх розв'язків системи обмежень називається *областю існування розв'язків*. Опорний розв'язок, при якому цільова функція набуває екстремального значення, називається *оптимальним*.

Детальніше алгоритми розв'язування задач лінійного програмування вивчаються в курсі математичного програмування, методів оптимізації та дослідження операцій.

Задача планування виробництва. Класичним ***прикладом*** найпростішої задачі системного аналізу в умовах визначеності є задача про виробництво і поставки товару. Деяка фірма повинна виробляти і постачати продукцію клієнтам однаковими партіями загальною кількістю $N = 24000$ одиниць на рік. Зрив поставок неприпустимий, оскільки штраф за це дуже великий. Згідно з умовами технології запускати у виробництво доводиться відразу всю партію. Вартість зберігання одиниці продукції на складі $C_x = 10$ гривень на місяць. Вартість запуску однієї партії у виробництво (незалежно від обсягу) $C_p = 40000$ гривень. Таким чином, запускати в рік багато партій явно невигідно, бо будуть великі витрати на запуск партій. Але невигідно і випускати мало партій, бо тоді обсяг партій стане дуже великим і витрати на зберігання сильно зростуть. Необхідно знайти «золоту середину» – визначити, яку кількість партій в рік найвигідніше випускати.

Побудуємо модель системи. Позначимо через n розмір партії і знайдемо кількість партій p за рік: $p = \frac{N}{n} = \frac{24000}{n}$.

Інтервал часу між запуском партій становить $t = \frac{12}{p}$ (місяців), а середній запас виробів на складі – $\frac{n}{2}$ штук. Випуск p партії по n штук буде коштувати $40000 \cdot p$ гривень. Складські витрати будуть коштувати $10 \cdot 12 \cdot \frac{n}{2}$ гривень. Отже, загальні річні витрати складають

$$E = C_x \cdot T \cdot \frac{n}{2} + C_p \cdot \frac{N}{n} \to min, \qquad (11.5)$$

де T – тривалість року в місяцях ($T = 12$).

Необхідно знайти таке n_0, при якому сума E досягає мінімуму. Розв'язання цієї задачі просте – треба взяти похідну від правої частини (11.5) по n і прирівняти її до нуля. Отримаємо

$$n_0 = \sqrt{\frac{2C_p \cdot N}{C_x \cdot T}}. \qquad (11.6)$$

Для нашого прикладу: $n_0 = 4000$ одиниць в одній партії. Це означає, що за рік треба випустити 6 партій і інтервал випуску між партіями становить 2 місяці. При цьому, витрати фірми (мінімальне значення цільової функції (11.5)) становлять $E = 4800$ гривень на рік.

Зіставимо цю суму з витратами при випуску 2000 виробів в партії або випуску партії один раз на місяць

$$E_1 = 0{,}1 \cdot 12 \cdot \frac{2000}{2} + 400 \cdot \frac{24000}{2000} = 6000 \text{ гривень на рік.}$$

А тепер візьмемо розмір партії 8000 виробів. Витрати становлять $E_2 = 0{,}1 \cdot 12 \cdot \dfrac{8000}{2} + 400 \cdot \dfrac{24000}{8000} = 6000$ гривень на рік. Наведені приклади підтверджують те, що навмання вибраний план зазвичай не є оптимальним.

Задачі управління запасами. Перші задачі управління запасами були описані ще в 1915 році – задовго не тільки до появи комп'ютерів, але й до вживання терміну «кібернетика» чи «програмування». Саме тоді було обґрунтовано метод розв'язання задачі – мінімізації витрат на замовлення та зберігання запасів при заданому попиті на дану продукцію та фіксованому рівні цін.

Трохи згодом були побудовані алгоритми розв'язання задачі управління запасами за більш складних умов – при змінному рівні цін (наявність «знижок за якість» і / або «знижок за кількість»); з урахуванням обмежень на складські потужності тощо.

11.3. Багатокритеріальні задачі. Критерії вибору альтернатив

Часто виникає ситуація, коли етап змістовної постановки задачі приводить до висновку про наявність декількох основних цілей функціонування системи. Справді, якщо деяка економічна система може мати «головну мету» – досягнення максимального прибутку, то майже завжди спостерігається наявність кількох часто суперечливих підцілей (збільшення/зменшення випуску продукції, розширення/згортання асортименту, завоювання ринку, збільшення/ зменшення ціни тощо), а також обмежень (фінансових, ресурсних) або умов (політична ситуація, погодні чи кліматичні умови тощо). Порушення цих обмежень і умов є неможливим чи катастрофічним (руйнується цілісність системи), або ж неприпустимими (є наслідки для зовнішнього середовища).

Розглянемо найпростішу ситуацію багатокритеріальності, за якої існують лише дві мети системи T_1 та T_2 і тільки дві можливі стратегії S_1 та S_2. Нехай оцінка ефективності стратегії S_1 стосовно мети T_1 становить $E_{11} = 0{,}4$ (за деякою шкалою $0 \div 1$). Отримавши таку ж оцінку для всіх пар «стратегія – ціль», можна побудувати таблицю – матрицю ефективності (табл. 11.1)

Таблиця 11.1

Матриця ефективності E стратегій S_1, S_2 стосовно цілей T_1, T_2

	T_1	T_2
S_1	0,4	0,6
S_2	0,7	0,3

Яку ж із стратегій вважати найкращою? Поки не буде визначено значимість кожної з цілей і не вказано їхні ваги – сперечатися марно! От якби було відомо, що перша мета, наприклад, у 3 рази важливіша за другу, то тоді можна оцінити їх відносні ваги – скажімо, величинами $0{,}75$ для першої та $0{,}25$ для другої. За таких умов сумарні ефективності стратегій (стосовно всіх цілей) складуть:

- для S_1 – $E_1 = 0{,}4 \cdot 0{,}75 + 0{,}6 \cdot 0{,}25 = 0{,}3 + 0{,}15 = 0{,}45$;
- для S_2 – $E_2 = 0{,}8 \cdot 0{,}75 + 0{,}2 \cdot 0{,}25 = 0{,}6 + 0{,}05 = 0{,}65$.

Отже, критерій ефективності системи (E_s) за наявності кількох цілей можна виразити через ефективності окремих стратегій (S_t) у вигляді

$$E_s = \sum S_t W_t , \qquad (11.7)$$

де W_t – ваги окремих цілей.

Рідко вдається встановити вагові коефіцієнти цілей за «фізичним змістом» задачі. Часто вагові коефіцієнти просто призначаються шляхом голосування або в інший спосіб. Справа в тому, що в ситуаціях, коли немає числового методу оцінки ваги мети, реальним виходом є використання накопиченого досвіду. Нерідко вагові коефіцієнти задає безпосередньо ОПР, але частіше його досвід управління підказує: «одна голова – добре, а декілька – краще», і ОПР приймає особливе рішення – використати метод експертних оцінок.

Суть цього методу досить проста. Потрібно чітко обумовити всі цілі функціонування системи і запропонувати групі осіб, висококомпетентних в даній галузі (експертів), розташувати всі цілі за значущістю або, як кажуть, за рангом.

Вищий ранг (зазвичай, 1) означає найбільшу важливість (вагу) цілі, наступний за ним – дещо меншу вагу тощо і так до 0. Спеціальні методи рангової кореляції дозволяють встановити узгодженість думок експертів або рангову конкордацію. Це особливо важливо у випадках, коли є потреба використовувати думки експертів, але існує сумнів у їх компетентності.

Критерії вибору альтернатив

Часто трапляються ситуації, коли особа, яка приймає рішення, може реалізувати тільки одну дію з декількох можливих, після чого уже неможливо повернутися до попереднього стану системи. Можливі варіанти дій називають ***альтернативами***. Серед всіх альтернатив можна вибрати найкращу, якщо є спосіб порівняння альтернатив між собою (***критерій переваги***).

Позначимо x – деяку альтернативу з множини X. Вважають, що для всіх $x \in X$ можна задати функцію $q(x)$, яка називається ***критерієм переваги*** і має таку властивість: якщо альтернатива x_1 переважає альтернативу x_2, то $q(x_1) > q(x_2)$, і навпаки. Найкращою вважається альтернатива x^* – така, за якої критерій набуває свого найбільшого значення

$$x^* = \underset{x \in X}{arg\,max}\, q(x). \qquad (11.8)$$

Інколи для більш об’єктивної оцінки альтернатив необхідно використовувати декілька критеріїв. *Приклад*: вибір президента країни; критерії – освіта, досвід політичної діяльності, патріотизм. При застосуванні декількох критеріїв вибір найкращої альтернативи стає неоднозначним. Розглянемо способи прийняття рішень у багатокритеріальних задачах.

Перший спосіб вибору найкращої альтернативи – зведення багатокритеріальної задачі до однокритеріальної. Зазвичай критерії нерівнозначні між собою (одні з них важливіші, ніж інші). Можливий варіант узагальненого критерію – це адитивна функція

$$Q(x) = w_1 q_1 + w_2 q_2 + \ldots + w_n q_n, \qquad (11.9)$$

де w_i – ваги окремих критеріїв.

Співвідношення (11.9) називають згорткою критеріїв (або, згенерованим критерієм).

Другий спосіб вибору найкращого рішення передбачає максимізацію головних критеріїв за умови, що менш важливі критерії не будуть знижуватися нижче за деякі встановлені межі.

Третій спосіб відшукання оптимальної альтернативи є можливим, якщо відомі оптимальні значення всіх критеріїв $q_1^0, q_2^0, \ldots, q_n^0$. Використовуючи поняття фазового простору і фазової точки поточного стану розглянемо відстань між поточним та оптимальним станом

$$d = \sqrt{\left(q_1 - q_1^0\right)^2 + \left(q_2 - q_2^0\right)^2 + \ldots + \left(q_n - q_n^0\right)^2}\,. \qquad (11.10)$$

Ідея оптимізації рішення полягає у побудові такої фазової траєкторії, кожна наступна точка якої буде ближчою до оптимальної точки $q_1^0, q_2^0, \ldots, q_n^0$.

Четвертий спосіб багатокритеріального вибору, який можна повністю формалізувати, полягає у відмові від єдиної «найкращої» альтернативи. Перевагу одній альтернативі перед іншою можна віддавати тільки тоді, коли перша за всіма критеріями краща, ніж друга. В іншому випадку альтернативи вважають непорівнянними. В результаті попарного порівняння альтернатив усі гірші за всіма критеріями альтернативи відкидають, а ті, що залишилися (непорівнянні між собою) – приймають. Прийняті альтернативи утворюють ***множину Парето***.

Вибір найкращої альтернативи можна описувати за допомогою ***мови бінарних відношень***. Апарат бінарних відношень у теорії прийняття рішень є теоретичним підґрунтям для оцінювання переваг альтернатив шляхом попарних порівнянь.

Бінарне відношення R на множині X визначається як деяка множина пар (x, y), де $x, y \in X$. Якщо пара (x, y) знаходиться у відношенні R, то це позначають xRy або $(x, y) \in R$. Існують різні способи подання бінарного відношення. Наприклад, матричне подання відношення R має вигляд

$$a_{ij}(R) = \begin{cases} 1, \textit{якщо } x_i \, R \, y_j, \\ 0, \textit{якщо } x_i \, \overline{R} \, y_j. \end{cases} \qquad (11.11)$$

Відомий *приклад* такого подання відношень – турнірні таблиці, в яких нічиї та програші позначено нулями, а виграші – одиницями. Прикладом графічного задання відношень є графи.

Бінарні відношення можуть мати такі властивості: рефлексивність чи антирефлексивність, симетричність, асиметричність чи антисиметричність, транзитивність чи антитранзитивність.

Приклади: відношення «$\le$» на множині чисел рефлексивне, антисиметричне, транзитивне; відношення «$<$» – антирефлексивне, асиметричне, транзитивне; відношення «мати спільний дільник» – рефлексивне, транзитивне; відношення «бути сином» на множині людей – антирефлексивне, антитранзитивне; відношення «бути сусідами» – симетричне, транзитивне.

Бінарне відношення, що має властивості рефлексивності, симетричності і транзитивності, називається ***відношенням еквівалентності*** і розбиває множину X на неперетинні ***класи еквівалентності***. *Приклади*: відношення «=» на множині чисел; відношення «вчитися в одному класі» розбиває множину учнів школи на окремі класи; відношення «мати однакове ім'я» розбиває множину людей на групи людей з однаковими іменами тощо.

11.4. Метод голосування

Один з найрозповсюдженіших методів прийняття рішень у соціальних групах – ***метод голосування***, який полягає в тому, що приймається альтернатива, котра одержала найбільшу кількість голосів. Цей спосіб приваблює своєю простотою та демократичністю, але, як і будь-який інший, має свої переваги та недоліки і застосовувати його треба обережно. До переваг, окрім вищесказаного, також можна віднести зростання поінформованості осіб, що беруть участь у голосуванні, оскільки перед голосуванням, як правило, проводиться відкрита дискусія щодо проблеми; окремі ОПР можуть змінити свої попередні висновки внаслідок такого обговорення. До недоліків методу голосування можна віднести відсутність анонімності при обговоренні питання, а значить і можливість неформального впливу більш авторитетних осіб внаслідок конформізму більшості, навіть за наявності протилежного власного висновку. Часто трапляється різна активність людей, яка в багатьох випадках не корелюється з їх компетентністю, адже не всі особи, які приймають участь у голосуванні, однаково обізнані з ситуацією.

Інколи правило більшості не спрацьовує: наприклад, при розподілі голосів порівно в разі парної кількості тих, хто голосував. Частина експертів взагалі може не проголосувати. Це породжує варіанти госування, наприклад: «керівник має два голоси», «відносна більшість (більше 51%)», «переважна більшість (більше $^2\!/_3$)», «принцип одностайності (консенсус, право вето)» тощо.

Розглянемо більш детально основні варіанти голосування.

Правило 1. (***абсолютної більшості***). Кожен виборець вибирає лише одну альтернативу. Перемагає та з них, яка набирає найбільшу кількість голосів. *Приклад*: вибори президента в Україні.

Правило 2. Перемагає альтернатива, яка переважає будь-яку іншу за правилом відносної більшості. Проте можлива така конфігурація переваг, за якої не буде переможця (***парадокс Кондорсе***). *Наприклад*: Порту – Шахтар 1:0, Шахтар – Зеніт 1:0, Зеніт – Порту 1:0.

Правило 3. (***де Борда***). Кожен виборець ранжує альтернативи від найкращої до найгіршої. Альтернатива отримує 1 бал за останнє місце, 2 бали – за передостаннє, …, n балів – за перше місце. Перемагає альтернатива з найбільшою сумою балів.

11.5. Експертні методи прийняття рішень

У процесі дослідження складних систем виникають проблеми, які важко описати математично. Тоді звертаються до послуг експертів – досвідчених спеціалістів, які можуть зменшити складність проблеми.

Особливий клас методів, що пов’язані безпосередньо з опитуванням експертів, утворюють ***методи експертних оцінок*** (назва походить від специфіки процедури отримання інформації від експертів – в анкетах експерти проставляють оцінки в балах чи рангах). Експертні методи використовуються в тих випадках, коли фахівці не лише не можуть відразу описати проблему за допомогою кількісних аналітичних залежностей, але і не бачать, які з методів формалізованого представлення систем могли б допомогти одержати модель для ухвалення рішення.

В залежності від конкретної задачі використовуються різні форми експертного опитування (різні види анкетування), підходи до оцінювання (ранжування, нормування, різні види упорядкування тощо), методи опрацювання результатів, вимоги до експертів, оцінювання їхньої компетентності, методики організації експертних опитувань.

В процесі опитування експертам роздають анкети з проханням оцінити пропоновані альтернативи. Експерти можуть оцінювати альтернативи у числовій або порядковій шкалі. Розглянемо перший випадок.

Нехай $q_j(x_i)$, $i=\overline{1,m}$, $j=\overline{1,n}$ – оцінка i-ї альтернативи j–м експертом,. За істинну характеристику цієї альтернативи вибирають вибіркове середнє:

$$\bar{q}(x_i)=\frac{1}{n}\sum_{j=1}^{n}q_j(x_i)\,. \qquad (11.12)$$

Узгодженість думок експертів вимірюють за допомогою величини

$$s_i=\frac{1}{n}\sum_{j=1}^{n}\left(q_j(x_i)-\bar{q}(x_i)\right)^2\,. \qquad (11.13)$$

Якщо експерти лише впорядковують альтернативи (порядкова шкала), арифметичні операції з оцінками є неможливими.

Розглянемо методи обробки експертної інформації у цьому випадку.

1. Результати опитування експертів зводять у таблицю, рядки якої відповідають альтернативам, стовпці – експертам. В i–му рядку записують місця (ранги), дані експертами i–ій альтернативі. В останній рядок записують суми рангів, отриманих об’єктами від експертів. Об’єкти впорядковують відповідно до сум рангів. На перше місце ставлять об’єкт, у якого вона найменша, і т. д.

Ступінь узгодженості думок експертів визначають за допомогою ***коефіцієнта конкордації***:

$$W = \frac{12\sum_{i=1}^{k} D_i^2}{k^2\left(n^3 - n\right)}. \qquad (11.14)$$

Тут

$$D_i = r_i - 0{,}5k\left(n+1\right), \qquad (11.15)$$

де n – кількість об’єктів; k – кількість експертів; r_i – сума рангів i –го об’єкта.

При опитуванні експертів може виникнути проблема низькокваліфікованого експерта.

Наприклад: розглянемо результати експертизи товарів табл. 11.2 ($n = 6$*;* $k = 4$).

Таблиця 11.2

Зразок	Експерт				Сума рангів			Ранг
	1	***2***	***3***	***4***	ΣR_{ij}	D_i	D_i^2	R
1. Тойота	1	2	1	5	9	-5	25	**1**
2. Фольксваген	2	1	3	4	10	-4	16	**2**
3. Шкода	4	5	3	6	18	4	16	**5**
4. Форд	5	4	6	1	16	2	4	**4**
5. Міцубісі	3	3	3	3	12	-2	4	**3**
6. Рено	6	6	5	2	19	5	25	**6**
S r	21	21	21	21	84	0	90	

Коефіцієнт конкордації у табл. 11.2 дуже низький:

$$W = \frac{12 \cdot 90}{16 \cdot 210} = 0{,}32\,.$$

Це свідчить про те, що думки експертів неузгоджені. Для оцінювання фахового рівня експертів додамо до табл. 11.2 ще один рядок «Оцінка експерта» (точніше кажучи, оцінка «розходження» думок експерта з іншими), значення в якому розрахуємо за формулою

$$b_j = \sum_{i=1}^{n} \left(a_{ij} - R_i \right)^2 , \qquad (11.16)$$

де a_{ij} – ранг i–го об'єкта на думку j–го експерта; R_i – ранг i–го об'єкта, узгоджений всіма експертами.

Зразок	Експерт				Сума рангів			Ранг
	1	***2***	***3***	***4***	ΣR_{ij}	D_i	D_i^2	R
1. …								
Оцінка експерта	**2**	**2**	**10**	**46**			0,32	

Чим менша оцінка, розрахована за формулою (11.16), тим вищий рівень експерта. Розглянувши оцінки експертів, відмічаємо низький фаховий рівень четвертого експерта. Виключаємо четвертого експерта з процедури голосування. Отримуємо нову табл. 11.3.

Таблиця 11.3

Зразок	Експерт			Сума рангів			Ранг
	1	***2***	***3***	ΣR_{ij}	D_i	D_i^2	
1. Тойота	1	2	1	4	-6.5	42.25	**1**
2. Фольксваген	2	1	3	6	-4.5	20.25	**2**
3. Шкода	4	5	3	12	1.5	2.25	**5**
4. Форд	5	4	6	15	4.5	20.25	**4**
5. Міцубісі	3	3	3	9	-1.5	2.25	**3**
6. Рено	6	6	5	17	6.5	42.25	**6**
S r	21	21	21	63	0	129.5	
Оцінка експерта	0	4	4			0.82	

Коефіцієнт конкордації у цьому випадку

$$W = \frac{12 \cdot 129{,}5}{216 - 6} = 0{,}82.$$

Високе значення коефіцієнта конкордації (близьке до 1) свідчить про високу узгодженість думок експертів. Як видно з табл. 11.3, перший експерт виявився найближчим до загального рішення. Це свідчить про його високу кваліфікацію.

На відміну від попереднього методу, інколи деяким об'єктам приписують однакові ранги, тобто вони «ділять між собою місця» (нестроге ранжування). Такі ранги називають зв'язаними, а їх група – зв'язкою. Для зв'язаних рангів здійснюють усереднення і кожному зв'язаному рангу приписують усереднене значення. ***Коефіцієнт конкордації для нестрогого ранжування*** визначають за формулою

$$W = \frac{12\sum_{i=1}^{k} D_i^2}{k^2\left(n^3 - n\right) - k\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(t_{ij}^3 - t_{ij}\right)}, \qquad (11.17)$$

де k_i – кількість груп однакових рангів за оцінкою i–го експерта; t_{ij} – кількість однакових рангів у j–й групі за оцінкою i–го експерта (у наведеному вище прикладі $k_i = 1, t_{i1} = 3$).

Наприклад 3, 4 і 6 об'єкти на думку 1–го експерта однакові, тоді кожному з них приписується ранг 5. Значення коефіцієнта конкордації у цьому випадку буде $W = 0{,}85$.

Інколи використовують інші підходи до оцінювання достовірності експертного рішення. При цьому визначають ступінь однодумності експертів за допомогою:

- середнього квадратичного відхилення;
- коефіцієнта варіації (C), який вираховується за формулою

$$C = \frac{\sigma \cdot 100\%}{M}, \tag{11.18}$$

де σ – середнє квадратичне відхилення оцінок експертів; M – середнє арифметичне значення.

Коефіцієнт варіації $> 30\%$ означає невдалий підбір експертної групи, високу неоднорідність її за ступенем компетентності щодо обговорюваної проблеми та неможливість вважати отриманий результат значимим [22].

2. **Метод парних порівнянь**. Часто виникає ситуація, коли експерту важко відразу оцінити якість великої кількості альтернатив і вибрати найкращу чи дати їм рангові оцінки. Метод парних порівнянь дозволяє уникнути складних порівнянь, замінивши їх попарними порівняннями, коли порівнюються лише два об'єкти шляхом визначення переваги одного над іншим.

Метод парних порівнянь полягає в тому, що на підставі зазначених l–им експертом переваг будують матрицю A^l, елементи якої визначаються таким чином

$$a_{ij}^l = \begin{cases} 1, \; \textit{якщо } i \succ j \textit{ для } l - \textit{го експерта} \\ 0, \; \textit{у протилежному випадку} \end{cases} \tag{11.19}$$

Розглянемо приклад таблиці, побудованої одним з експертів (табл. 11.4).

Таблиця 11.4

Таблиця парних порівнянь

	Тойота	Фольксваген	Шкода	Форд	Міцубісі	Рено	Сума
Тойота		1	1	1	1	1	**5**
Фольксваген	0		1	1	1	1	**4**
Шкода	0	0		1	0	1	**2**
Форд	0	0	0		0	1	**1**
Міцубісі	0	0	1	1		1	**3**
Рено	0	0	0	0	0		**0**

Обчислюють суми рядків $a_i = \sum_{j=1}^{k} a_{ij}$. Альтернативи впорядковують за спаданням величин a_i. Альтернатива із найбільшим значенням a_i є переважаючою.

В методі парних порівнянь окрім двобальної, можуть використовуватися і інші шкали відносної важливості, наприклад, дев'ятибальна (табл. 11.5).

Таблиця 11.5

Дев'ятибальна шкала відносної важливості

Бал	Визначення
1	Рівна важливість
3	Помірна перевага
5	Суттєва перевага
7	Значна перевага
9	Дуже велика перевага
2, 4, 6, 8	Проміжні значення

Наприклад, порівнюючи чотири моделі комп'ютерів (А, Б, В, Г) експерт дав такі оцінки у вигляді тверджень:

- модель В помірно переважає модель А;
- модель Б помірно поступається моделі В;
- модель А значно переважає модель Г;
- перевага моделі В над Г знаходиться між значною та дуже великою;
- модель А помірно поступається моделі Б;
- модель Б помірно переважає модель Г.

Відповідна матриця парних порівнянь матиме вигляд

	А	Б	В	Г	Сума
А	1	3	1/3	7	11,3
Б	1/3	1	1/3	3	4,7
В	3	3	1	8	15,0
Г	1/7	1/3	1/8	1	1,6

Очевидно, що модель В за таких експертних оцінок має найбільшу перевагу над іншими.

Теми для роздумів і контрольні запитання

1. Що таке управління системою? Яка мета управління системою?
2. Як часто Ви приймаєте рішення?
3. У чому полягають труднощі прийняття рішень?
4. Які етапи включає в себе процедура прийняття рішення в загальному випадку?
5. Яка відмінність між особами, що приймають рішення (ОПР), експертами та консультантами (аналітиками)?
6. У чому полягає проблема узгодження цілей? Наведіть приклади.
7. Поняття задачі математичного програмування.
8. Що таке багатокритеріальні задачі прийняття рішень?
9. Які існують підходи до моделювання багатокритеріальних задач прийняття рішень?

10. Поняття критерію переваги. Способи прийняття рішень у багатокритеріальних задачах.
11. Що називають множиною Парето?
12. Поняття бінарного відношення. Властивості бінарних відношень.
13. Метод голосування як метод прийняття рішень. Переваги і недоліки.
14. В чому полягає суть методів експертних оцінок?
15. Яким чином вимірюють узгодженість думок експертів?
16. Які є методи обробки експертної інформації?
17. Що таке коефіцієнт конкордації?
18. Як визначається коефіцієнт конкордації для нестрогого ранжування?
19. В чому полягає суть методу парних порівнянь?
20. Шкали відносної важливості.

12. ПРИЙНЯТТЯ РІШЕНЬ В УМОВАХ НЕВИЗНАЧЕНОСТІ

12.1. Схема матричної гри з «природою»

Невизначеність виникає у завданнях прийняття рішень, у яких ОПР не знає всієї сукупності чинників, що діють, або не знає закономірностей їх прояву. Ситуація невизначеності характеризується тим, що вибраний ОПР план дій повинен бути реалізованим при певній сукупності зовнішніх чинників, яка випадковим чином реалізується з певної множини варіантів. Класична задача прийняття рішень за умов невизначеності полягає в тому, що ОПР знає можливі варіанти своїх дій, можливі варіанти реалізації зовнішніх чинників та значення очікуваного виграшу (програшу), який відповідає комбінації «рішення – чинники», але не знає детермінованого механізму прояву зовнішніх чинників.

Нехай ОПР може обрати одну із можливих альтернатив-стратегій, а зовнішнє середовище (природа) у відповідь, «не зважаючи» на вибір ОПР, обирає один зі своїх станів. Під терміном «природа» тут розуміють вплив на ситуацію сукупності чинників, які «непідвладні» ОПР і проявляються випадковим чином. Зокрема, «природою» може бути зовнішнє середовище, дії ОПР фірми–конкурента чи політична ситуація в країні.

Таку ситуацію прийняття рішення відображає матриця виграшів F, рядки якої відповідають вибору ОПР однієї зі своїх стратегій $S_i, i = 1, \ldots, n$ (свідомий вибір), а стовпчики – вибору «природою» одного зі своїх можливих станів C_j, $j = 1, \ldots, m$ (випадковий, байдужий до дій ОПР, вибір «природи»). Елементами матриці F є числа F_{ij} – виграші ОПР (якщо значення F_{ij} від'ємне – програші ОПР).

Як *приклад* моделі матричної гри з «природою», розглянемо взаємодію людини з навколишнім середовищем. В ролі навколишнього середовища будемо розглядати стан природного середовища. Вважатимемо, що навколишнє середовище (клімат) може перебувати в наступних станах: C_1 – дуже сухий, C_2 – сухий, C_3 – помірний, C_4 – вологий, C_5 – дуже вологий. Стратегія особи, яка приймає рішення (фермера), полягає у виборі розміру інвестицій у зрошення сільськогосподарських культур: S_1 – кошти не інвестуються; S_2 – інвестується 50% планових коштів; S_3 – повна планова інвестиція. Вибираючи одну із можливих стратегій, ОПР отримує в результаті деякий фінансовий виграш (програш) $F_{ij} = F(S_i, C_j)$, величина якого залежить від стану зовнішнього середовища.

Наприклад, якщо фермер прийме рішення про максимальні інвестиції, а сезон буде вологим, це означатиме даремно витрачені кошти. Навпаки, якщо фермер вирішить не фінансувати зрошення, а сезон буде сухим, розмір врожаю буде нижчим від очікуваного.

Таким чином, матриця F_{ij} характеризує виграш гравця у разі, якщо він вибере рішення S_i, а середовище перебуватиме у стані C_j

$$F = \begin{array}{cc} \begin{array}{cccc} C_1 & C_2 & \ldots & C_m \end{array} & \\ \begin{pmatrix} F_{11} & F_{12} & \ldots & F_{1m} \\ F_{21} & F_{22} & \ldots & F_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ F_{n1} & F_{n2} & \ldots & F_{nm} \end{pmatrix} & \begin{array}{c} S_1 \\ S_2 \\ \ldots \\ S_n \end{array} \end{array}.$$

У залежності від ситуації, інформація про ймовірності станів «природи» може бути відомою, або ж невідомою. Мета моделювання такої задачі прийняття рішення – надати рекомендації ОПР щодо вибору оптимальної стратегії.

Оптимальна стратегія гравця в умовах невизначеності обирається на основі зведення стохастичної задачі прийняття рішень до детермінованої задачі на оптимізацію. Відповідно до кількості наявної інформації про навколишнє середовище вибирають ті чи інші критерії прийняття рішень.

Зауважимо, що на прийняття рішення впливають також і психологічні чинники, зокрема, відношення до ризику особи, яка приймає рішення (обережна стратегія, ризикована стратегія, зважена стратегія). З урахуванням індивідуальних особливостей виділяють обережних керівників, які дотримуються стратегії мінімізації очікуваних втрат, та ризикових керівників, які більше схильні до рішень з очікуваним максимальним виграшем.

12.2. Критерії прийняття рішень в умовах невизначеності

Розробка рішень в умовах невизначеності здійснюється за допомогою математичних моделей теорії ігор. При цьому стосовно конкретних ситуацій вибір дії визначається не тільки оцінкою результату, а й можливими альтернативами дій конкурентів або впливами інших елементів навколишнього середовища.

Перша інформаційна ситуація характеризується тим, що відомі ймовірності станів навколишнього середовища і вони не залежать від часу. Вважається, що рішення може прийматися (теоретично) нескінченну кількість разів. За таких умов критеріями прийняття рішень можуть бути критерій Байєса чи критерій мінімальної дисперсії.

Критерії Байєса (середнього ризику). Вважається, що ймовірності станів навколишнього середовища розраховуються на підставі тривалих спостережень за станом «природи».

Повертаючись до вищенаведеного *прикладу*, позначимо через p_1 – ймовірність настання дуже сухого стану довкілля, p_2 – ймовірність настання сухого стану, p_3 – ймовірність настання помірного стану, p_4 – ймовірність настання вологого стану, p_5 – ймовірність настання дуже вологого стану. Має виконуватися умова: $p_1 + p_2 + p_3 + p_4 + p_5 = 1$ (події утворюють повну групу).

Критерій Байєса оцінює стратегії за максимальним значенням середнього за імовірністю виграшу

$$F_{Bi} = \sum_{j=1}^{N} F_{ij} \cdot p_j \to max\,, \qquad (12.1)$$

де N – кількість можливих станів середовища (у нашому випадку $N = 5$).

Нехай конкретна матриця виграшів має наступний вигляд (табл 12.1).

Таблиця 12.1

Матриця 1 виграшів та ймовірності станів «природи»

Стратегії гравця	Стани навколишнього середовища				
	дуже сухий	сухий	помірний	вологий	дуже вологий
S_1	50	70	110	130	110
S_2	80	110	120	90	60
S_3	100	130	110	80	60
Ймовірності станів ***P***	0.10	0.30	0.30	0.20	0.10

Розраховуємо значення критерію Байєса (середньозважений виграш) для кожної із стратегій за формулою (12.1)

$$F_{B1} = 50 \cdot 0{,}10 + 70 \cdot 0{,}30 + 110 \cdot 0{,}30 + 130 \cdot 0{,}20 + 110 \cdot 0{,}10 = 96\,;$$
$$F_{B2} = 80 \cdot 0{,}10 + 110 \cdot 0{,}30 + 120 \cdot 0{,}30 + 90 \cdot 0{,}20 + 60 \cdot 0{,}10 = 101\,;$$
$$F_{B3} = 100 \cdot 0{,}10 + 130 \cdot 0{,}30 + 110 \cdot 0{,}30 + 80 \cdot 0{,}20 + 60 \cdot 0{,}10 = 104\,.$$

Згідно з критерієм Байєса оптимальною стратегією у цьому випадку є стратегія 3 (повна планова інвестиція), яка відповідає максимальному очікуваному виграшу.

Проте, критерій Байєса не завжди дозволяє визначити оптимальну стратегію (див., напр., табл. 12.2). Як бачимо, критерій Байєса дає однакову оцінку обох стратегій. В такому випадку варто додатково порівняти ступінь ризику кожної із стратегій, оскільки багатьох керівників відлякують ситуації, у яких ступінь ризику є високим. Для оцінки ступеня ризику використовують дисперсію – середньостатистичне відхилення можливих виграшів від середнього виграшу.

Таблиця 12.2

Матриця 2 виграшів та ймовірності станів «природи»

Стратегії гравця	Стани навколишнього середовища			Критерій Байєса	Дисперсія
	сухий	помірний	вологий		
S_1	50	100	150	100	1630
S_2	80	100	120	100	**132**
Ймовірності станів ***P***	0.33	0.34	0.33		

Оскільки перша стратегія є більш ризикованою, перевагу потрібно віддати другій.

Критерій мінімальної дисперсії відповідає типу поведінки – «максимальна обережність» і розраховується за мінімальним значенням виразу

$$F_{md\,i} = \sum_{j=1}^{N} \left(F_{ij} - F_{b_i}\right)^2 \cdot p_j \to min\,. \qquad (12.2)$$

Обчислимо значення критерію мінімальної дисперсії для кожної із стратегій за даними табл. 12.1.

$$F_{md1} = (50-96)^2 \cdot 0{,}10 + (70-96)^2 \cdot 0{,}30 + (110-96)^2 \cdot 0{,}30 +$$
$$+ (130-96)^2 \cdot 0{,}20 + (110-96)^2 \cdot 0{,}10 = 724;$$
$$F_{md2} = (80-101)^2 \cdot 0{,}10 + (110-101)^2 \cdot 0{,}30 + (120-101)^2 \cdot 0{,}30 +$$
$$+ (90-101)^2 \cdot 0{,}20 + (60-101)^2 \cdot 0{,}10 = 369;$$
$$F_{md3} = (100-104)^2 \cdot 0{,}10 + (130-104)^2 \cdot 0{,}30 + (110-104)^2 \cdot 0{,}30 +$$
$$+ (80-104)^2 \cdot 0{,}20 + (60-104)^2 \cdot 0{,}10 = 524.$$

Згідно з критерієм мінімальної дисперсії оптимальною є стратегія 2 (інвестувати 50% планових коштів). Така стратегія є найбільш обережною.

Друга інформаційна ситуація характеризується тим, що розподіл ймовірностей станів навколишнього середовища є невідомим. Якщо в сільськогосподарському виробництві, що залежить від погодних умов, можлива орієнтація на середній ризик, то в деяких інших ситуаціях, які загрожують великими втратами, доцільно обирати найобережнішу стратегію. Такими, наприклад, є задачі про розрахунок ризику повені, землетрусу, аварії атомної електростанції.

Критерій Вальда. У таких задачах, де «природа» виступає агресивним і підступним суперником, а неправильний вибір стратегії загрожує значними втратами, для вибору стратегії поведінки ОПР використовують ***максимінний*** критерій (критерій Вальда), орієнтований на найнесприятливіші умови середовища

$$Fw_i = \min_j F_{ij} \to max\,. \qquad (12.3)$$

Розрахуємо значення критерію Вальда для кожної із стратегій за даними табл. 12.1:

$$F_{W1} = \min(50,\ 70, 110, 130, 110) = 50\,;$$

$$F_{W2} = \min(80, 110, 120, 90, 60) = 60\,;$$
$$F_{W3} = \min(100, 130, 110, 80, 60) = 60\,.$$

Згідно з критерієм Вальда оптимальними є стратегія 2 або стратегія 3.

Критерій Севіджа. Критерій Севіджа є антиподом до критерію Вальда. Якщо критерій Вальда орієнтується на найнесприятливіші стани середовища, то критерій Севіджа орієнтується на найсприятливіші стани середовища. Тому його часто називають оптимістичним критерієм або ***максимаксним*** критерієм

$$Fs_i = \max_j F_{ij} \to \max\,. \tag{12.4}$$

Для розрахунку критерію Севіджа остання колонка таблиці заповнюється максимальними значеннями кожного рядка матриці виграшів. Потім з них вибирається максимальне. Згідно з таким критерієм, у випадку табл 12.2 оптимальною є перша стратегія, оскільки вона дозволяє отримати максимальний виграш.

Обчислимо значення критерію Севіджа для кожної із стратегій (табл 12.1)

$$F_{S1} = \max(50, 70, 110, 130, 110) = 130\,;$$
$$F_{S2} = \max(80, 110, 120, 90, 60) = 120\,;$$
$$F_{S3} = \max(100, 130, 110, 80, 60) = 130\,.$$

Згідно з критерієм Севіджа оптимальними є стратегія 1 або стратегія 3.

Критерій Гурвіца. Припустимо тепер, що ОПР має певну інформацію про співвідношення оптимізму і песимізму при прийнятті рішення. Маємо **інформаційну ситуацію третього типу**, що є проміжною між першою і другою інформаційними ситуаціями. Суть критерію Гурвіца полягає у визначенні зваженої комбінації розв'язків Вальда і Севіджа. Використується зважуючий коефіцієнт λ, ($0 < \lambda < 1$), так званий, коефіцієнт песимізму. Коли $\lambda = 1$, отримаємо песимістичний критерій (критерій Вальда), а в разі $\lambda = 0$ – крайній оптимізм (критерій Севіджа). Критерій Гурвіца має вигляд

$$F_{H_i} = \lambda \min_j F_{ij} + (1 - \lambda) \max_j F_{ij} \to max. \qquad (12.5)$$

Виберемо значення $\lambda = 0{,}5$ і розрахуємо критерій Гурвіца для кожної із стратегій за даними табл. 12.1

$$F_{H_1} = 0{,}5 \cdot \min(50, 70, 110, 130, 110) +$$
$$+ 0{,}5 \cdot \max(50, 70, 110, 130, 110) = 90;$$
$$F_{H_2} = 0{,}5 \cdot \min(80, 110, 120, 90, 60) +$$
$$+ 0{,}5 \cdot \max(80, 110, 120, 90, 60) = 90;$$
$$F_{H_3} = 0{,}5 \cdot min(100, 130, 110, 80, 60) +$$
$$+ 0{,}5 \cdot max(100, 130, 110, 80, 60) = 95.$$

Згідно з критерієм Гурвіца оптимальною є стратегія 3.

Критерій Ходжа−Лемана найчастіше застосовують у випадках, коли ймовірності застосування «природою» своїх стратегій, загалом, невідомі, але можливі деякі припущення про розподіли ймовірності. Він ґрунтується на максимінному критерії та критерії Байєса. У критерії Ходжа−Лемана використовується параметр υ ($0 < \upsilon < 1$), що відповідає довірі до використовуваного розподілу ймовірностей. Якщо ступінь довіри великий, то застосовують критерій Байєса, а ні – то переважає складова, що відповідає максимінному критерію. Критерій Ходжа−Лемана має вигляд

$$F_{HLi} = \upsilon \sum_{j=1}^{N} F_{ij} \cdot p_j + (1-\upsilon) \min_j F_{ij} \to \max . \qquad (12.6)$$

У випадку $\upsilon = 1$ отримуємо критерій Байєса, а для $\upsilon = 0$ критерій максиміну.

Припустимо, що у нашому прикладі (табл. 12.1) є сумніви щодо ймовірностей стану економіки, і тому покладемо $\upsilon = 0{,}5$. За критерієм Ходжа−Лемана отримаємо

$$F_{HL1} = 0{,}5 F_{B1} + 0{,}5 F_{W1} = 98 ;$$
$$F_{HL2} = 0{,}5 F_{B2} + 0{,}5 F_{W2} = 110{,}5 ;$$
$$F_{HL3} = 0{,}5 F_{B3} + 0{,}5 F_{W3} = 112 .$$

Згідно з критерієм Ходжа−Лемана за відповідних припущень оптимальною стратегією є стратегія 3.

Критерій Бернуллі–Лапласа. В умовах повного незнання про поведінку середовища (**інформаційна ситуація четвертого типу** або «досконала» невизначеність) доцільно використати критерій Бернуллі–Лапласа. Суть його зводиться до того, що оскільки немає підстав віддавати перевагу одному із можливих сценаріїв, то всі вони вважаються рівноможливими. Тоді за математичне сподівання виграшу при стратегії S_i беруть середнє арифметичне виграшів при всіх варіантах стану середовища. Потім з них вибирають максимальне

$$F_{BLi} = \frac{1}{N} \sum_{j=1}^{N} F_{ij} \to max . \qquad (12.7)$$

Розраховуємо значення критерію Бернуллі–Лапласа для кожної із стратегій (табл. 12.1)

$$F_{BL1} = \frac{50 + 70 + 110 + 130 + 110}{5} = 94 ;$$

$$F_{BL2} = \frac{80+110+120+90+60}{5} = 92\,;$$

$$F_{BL3} = \frac{100+130+110+80+60}{5} = 96\,.$$

Згідно з критерієм Бернуллі–Лапласа оптимальною стратегією є стратегія 3.

Теми для роздумів і контрольні запитання

1. У чому полягає класична задача прийняття рішень за умов невизначеності?
2. Поняття матричної гри з «природою».
3. В чому відмінність інформаційних ситуацій першого, другого, третього та четвертого типу у математичних моделях теорії ігор?
4. Які критерії прийняття рішень в умовах невизначеності застосовуються у випадку першої інформаційної ситуації?
5. Порівняйте критерії Байєса і мінімальної дисперсії.
6. Які критерії прийняття рішень в умовах невизначеності застосовуються у випадку інформаційної ситуації другого типу?
7. Порівняйте критерії Вальда і Севіджа.
8. Які критерії прийняття рішень в умовах невизначеності застосовуються у випадку інформаційної ситуації третього типу?
9. В чому полягає суть критеріїв Гурвіца і Ходжа−Лемана?
10. Який критерій прийняття рішень застосовують у випадку повної невизначеності?

13. УПРАВЛІННЯ СИСТЕМОЮ

13.1. Основні принципи управління

Під управлінням розуміють здійснення сукупності дій, спрямованих на підтримку та покращення функціонування об'єкта управління. Основною вимогою до управління системою – є забезпечити її целеспрямовану поведінку при змінних зовнішніх умовах. Це досягається правильною організацією системи, під якою розуміють її структуру та спосіб функціонування.

Якщо організація системи однозначно визначена при її створенні, то управління зводиться до підтримання і забезпечення розрахункових значень параметрів системи, а також контролю їх відхилень та реакції на такі відхилення.

При формуванні системи ті її елементи, якими управляють, об'єднуються в підсистему – ***об'єкт управління*** (ОУ). Сукупність елементів, які здійснюють управління, також утворюють систему – ***управляючу систему*** (УС). Обидві підсистеми взаємодіють за допомогою інформаційних та організаційних зв'язків. Управління поведінкою системи здійснюють шляхом управління її входами X або ж незалежними від входів параметрами системи Q. Управління системою спрямоване на обмеження ступенів свободи та звуження діапазонів вихідних змінних.

Процес управління, зазвичай, складається з двох етапів:

1) розробка програми правильної поведінки ОУ;
2) реалізація програми: регулювання, керівництво, управління.

Структура управляючої системи будується за ***ієрархічним принципом***.

♦ Кожен рівень УС управляє нижчим рівнем і управляється вищим рівнем.

♦ Інформація рухається від нижніх рівнів до верхніх і при цьому послідовно стискується. Нижчий рівень управління постає перед вищим як «чорна скринька», яка інформує лише про результати діяльності, але нічого не повідомляє про внутрішні процеси, які відбуваються на цьому рівні.

♦ Чим вища самостійність даного рівня управління, тим вищий ступінь стискування інформації. Відносна самостійність кожного рівня і послідовне стискання інформації – головні умови ефективності багаторівневого управління.

♦ Ієрархія управляючої системи будується у відповідності до ієрархії цілей, які стоять перед системою.

Розрізняють три типи *основних завдань регулювання*: стабілізація, програмне регулювання та спостереження.

Мета ***стабілізації*** – підтримка постійного значення деякої вихідної величини.

Програмне регулювання забезпечує зміну вихідної змінної відповідно до заданої програми. Найчастіше його метою є стабілізація вихідного показника в околі заданого значення. Для цього найчастіше використовують принцип зворотного зв'язку: деяка залежність вхідного показника від значення вихідного.

Спостереження відрізняється тим, що управління здійснюється не згідно з наперед розробленою програмою, а в залежності від поведінки об'єкта. *Прикладом* є автоматичне підстроювання радіоприймача, що стежить за обраною хвилею.

Наявності зворотного зв'язку не завжди достатньо для забезпечення стійкості управління. Запізнювання, інерція системи, приховані нелінійності не можуть належним чином бути враховані при виборі параметрів зворотного зв'язку. Тоді найефективнішим регулюванням стає ***пристосування*** до характеристик середовища та об'єкта управління. Пристосування досягається в процесі навчання. Під ***навчанням*** розуміють накопичення інформації про хід процесу регулювання в минулому та її використання для удосконалення цього процесу на основі деякого набору правил.

Системи, які мають і програмний блок, і регулятор називають ***системами, що самоорганізуються***.

Управління системою завжди спрямоване на обмеження кількості ступенів свободи системи, або на обмеження діапазонів зміни її параметрів. Таким чином зменшується кількість варіантів поведінки системи (відкидаються небажані варіанти). В цьому полягає один з основних методів управління.

Для ефективного вирішення завдань управління необхідно, щоб інформаційна потужність керуючої підсистеми була вищою від потужності об'єкта управління. При цьому працює закон необхідної різноманітності Ешбі: різноманітність керуючої підсистеми повинна бути вищою від різноманітності об'єкта управління.

13.2. Керованість і спостережуваність систем

Розглянемо проблему керованості і спостережуваності лінійної динамічної системи. Для описання цих понять будемо використовувати модель фазового простору (простору станів системи). Як відомо, стан системи описується точкою у фазовому просторі. Координатами точки фазового простору виступають основні параметри системи (*наприклад*, для автомобіля: швидкість, напрям руху, кількість пального тощо). Кількість таких параметрів визначається розмірністю моделі системи.

Нехай динамічна система описується вектором стану Q

$$Q = \{q_1, q_2, \ldots, q_k, q_{k+1}, q_{k+2}, \ldots, q_m\}, \qquad (13.1)$$

вектором входів $X = \{x_1, x_2, \ldots, x_n\}$ та вектором виходів $Y = \{q_1, q_2, \ldots, q_k\}$.

Якщо вхідні змінні впливають на всі параметри системи, така система називається ***керованою***.

Якщо вектор стану Q співпадає з вектором виходів Y, тобто, всі параметри системи можна взнати (спостерігати) за вектором стану Q, то система називається ***спостережуваною***.

Значення параметрів $q_{k+1}, q_{k+2}, \ldots, q_m$ вектора стану Q, які неможливо взнати за вектором виходів Y, називають неспостережуваними параметрами. Якщо система містить неспостережувані параметри вона називається ***неспостережуваною***.

Нехай всі параметри системи типу (13.1) є спостережуваними, але вхідні параметри впливають лише на k перших параметрів стану $Y = \{q_1, q_2, \ldots, q_k\}$. Така система називається спостережуваною, але ***некерованою***.

Отже, всі системи можна розділити на 4 категорії:

- спостережувані, але некеровані;
- неспостережувані, але керовані;
- спостережувані і керовані;
- неспостережувані і некеровані.

Множина послідовних спостережуваних станів описує фазову траєкторію системи.

Аналітичним способом управління системою можна описати таким чином

$$\begin{cases} \dfrac{dx}{dt} = Ax + Bu, \\ y = Cx, \end{cases} \tag{13.2}$$

де $x = \{x_1, x_2, \ldots, x_n\}$ – вектор вхідних даних;

$y = \{y_1, y_2, \ldots, y_q\}$ – вектор вихідних параметрів;

$u = \{u_1, u_2, \ldots, u_m\}$ – вектор управління;

$A,\ B,\ C$ – дійсні постійні матриці розміру $n \times n, n \times q, n \times m$, які, наприклад, при $n = 2, \quad q = 2, \quad m = 3$ мають вигляд:

$$A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix};$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}; \qquad (13.3)$$

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}.$$

Якщо один з рядків матриці B (наприклад, перший) складається цілком з нульових елементів, тоді відповідна координата (перша) буде некерованою, оскільки жодна з управляючих дій не здатна чинити управляючий вплив на цю координату.

Аналогічно, якщо один із стовпців матриці C складається з нульових елементів, то відповідна координата вектора стану не чинить впливу ні на один з двох сигналів, що спостерігаються – (y_1, y_2). Її поведінка ніяк не буде проявлятися зовні, оскільки обидві координати є не спостережуваними.

У загальному випадку недіагональної матриці A умови керованості та спостережуваності є складнішими. Систему вважають керованою, якщо можна організувати такі входи у систему, щоб отримати потрібні виходи.

Система (13.2) називається **керованою**, якщо існує єдиний вектор u, який переводить цю систему із початкового стану x_0 у кінцевий стан x_n.

Достатньою умовою повної керованості системи є критерій Калмана.

Критерій Калмана. Лінійна стаціонарна система (13.2) є повністю керованою, якщо ранг матриці $\left[B, AB, A^2B, \ldots, A^{m-1}B\right]$ дорівнює m.

Систему вважають **спостережуваною**, якщо за відомими значеннями входів і виходів системи можна оцінити її стани.

Лінійна стаціонарна система (13.2) є цілком спостережуваною, якщо ранг матриці $\left[C^T, C^T A^T, C^T (A^T)^2, \dots, C^T (A^T)^{m-1}\right]$ дорівнює m.

Система називається ***структурно стійкою***, якщо вона зберігає тенденцію прагнення до того стану, який найкраще відповідає меті системи, цілям збереження якості без зміни структури або таким, що не приводять до суттєвих змін структури системи на деякій заданій множині ресурсів (наприклад, на часовому інтервалі). Нечітке поняття «суттєві зміни» має бути конкретизовано і детерміновано для кожної системи. Ілюстрацією до стійкості може бути маятник, який за будь-якого відхилення від стійкого стану «прагне» повернутися до стану рівноваги. Якщо розглядати лід як систему, то за температур, близьких до 0, ця система буде структурно нестійкою. Іншим прикладом структурно нестійкої системи може бути ринок за нестійкого співвідношення попиту–пропозиції.

Для аналізу і синтезу керованих систем необхідно вміти встановлювати їх ***асимптотичну стійкість***. Ця властивість показує здатність системи виходити на стабільний рівень функціонування після дії імпульсного впливу. *Прикладом* асимптотично стійкої системи є іграшка «Неваляшка». Для асимптотичної стійкості системи необхідно і достатньо, щоб всі власні числа матриці A знаходилися в одиничному колі комплексної площини, тобто $|\lambda_i| < 1, i = 1, \dots, n$.

13.3. Оптимізація управління системою

В процесі управління системою щоразу необхідно вибирати такий варіант дій, при якому можна досягти мети найкращим способом. Це завдання стає конкретним, якщо існує кількісна характеристика для результатів управління – показник ефективності управління. Такий показник називають ***цільовою функцією управління***.

Функціоналом в математиці називають функцію, областю визначення якої (аргументом) є множина деяких функцій, а множиною значень – дійсні числа.

Для оцінки якості управління найчастіше використовують інтегральний показник (функціонал) такого вигляду

$$I = \int_0^T f(x)dt\,, \tag{13.4}$$

де $f(x)$ – функція, яка залежить від параметрів стану системи.

Якщо $f(x)=1$, інтегральний показник оцінює час перехідного процесу (час, необхідний системі для переходу до оптимального стану)

$$I = \int_0^T dt\,. \tag{13.5}$$

Якщо $f(x)$ – функція, яка описує відхилення системи від оптимальних значень параметрів, інтегральний показник оцінюється за формулою

$$I = \int_0^T f^2(x)dt \to min\,. \tag{13.6}$$

Задача оптимального управління формулюється наступним чином.

Стан об’єкта управління описується вектором $x = \{x_1, x_2, \ldots, x_n\}$ – вектором фазових координат. Координати об’єкта змінюються в часі за законом

$$\frac{dx_i}{dt} = f_i(x,u), \quad \left(i = \overline{1,n}\right), \tag{13.7}$$

де $u = \{u_1, u_2, \dots, u_m\}$ – координати вектора управління.

Необхідно вибрати таке управління, для якого значення функціонала I є мінімальним

$$I = \int_{t_0}^{t_1} f_0(x(t), u(t))dt \to min\,. \tag{13.8}$$

В залежності від вигляду функції $f_0(x(t), u(t))$ функціонал (13.8) визначає ефективність управління за різними критеріями: критерієм часу, критерієм відстані, економічним критерієм тощо.

Принцип необхідної різноманітності вимагає відповідності між рівнем різноманітності (складності) управляючої системи і системи, якою управляють.

Зазвичай, складність об'єкта управління H_m^0 і системи управління H_m^S визначають за формулою Хартлі

$$H_m^0 = log\, n_0; \quad H_m^S = log\, n_S\,, \tag{13.9}$$

де n_0, n_S – кількість станів об'єкта і системи управління.

У цьому випадку критерій ефективного управління має вигляд

$$I = \int_0^T [log\, n_S - log\, n_0 - log\, a]^2\, dt \to min\,, \tag{13.10}$$

де a – потрібна різниця між порядками управляючої системи та об'єкта управління.

Важливим показником функціонування системи є її надійність. Під ***надійністю*** системи управління розуміють її здатність зберігати найбільш суттєві властивості на заданому рівні протягом визначеного проміжку часу. Надійність систем пов'язана з ймовірністю її відмов.

Відмова – подія, яка полягає у порушенні працездатності системи. Відмови поділяють на раптові і поступові. Основними показниками надійності системи є ймовірність безвідмовної роботи, інтенстивність (частота) відмов, середній час безвідмовної роботи тощо.

Теми для роздумів і контрольні запитання

1. Поняття організації управління системою. Об'єкт управління і управляюча система.
2. В чому полягає суть ієрархічного принципу управляючої системи?
3. Поняття регулювання, пристосування та навчання системи.
4. Яка система називається керованою, спостережуваною?
5. Достатня умова повної керованості системи.
6. Яка лінійна стаціонарна система є цілком спостережуваною?
7. Поняття структурної та асимптотичної стійкості системи. Наведіть приклади.
8. Що є показником ефективності управління системою?
9. Сформулюйте задачу оптимального управління системою.
10. Поняття надійності і відмовостійкості системи управління.

ГЛОСАРІЙ

Автокореляційна функція (**АКФ**) – послідовність коефіцієнтів автокореляції, які відповідають значенням лагу від 1 до k .

Автокореляція залишків – явище при якому спостерігається взаємозв’язок між послідовними значеннями залишків.

Атрактор – сукупність фазових траєкторій, які притягується до деякої скінченновимірної підмножини фазового простору.

Взаємна автономність – властивість елементів системи, яка проявляється в тому, що кожному її елементу притаманні властивості системи в цілому.

Випадковість – вид невизначеності, який підкоряється лише закону розподілу.

Вихід системи – канал впливу системи на зовнішнє середовище.

Відмова системи – подія, яка полягає у порушенні працездатності системи.

Вхід системи – канал, за допомогою яких зовнішнє середовище впливає на систему.

Генеральна сукупність – сукупність усіх можливих значень дискретної випадкової величини.

Граф – система, яка складається з двох множин: ***вершин*** (*vertices*) та ***ребер*** (*edges*), що їх з’єднують.

Дерево графу – зв’язний граф, який не має циклів.

Дивний атрактор – сукупність фазових траєкторій, які під дією невеликих змін початкових умов системи, характеризуються кардинальними змінами фазового портрету через певний проміжок часу.

Дуга графу – напрямлене ребро графу.

Еквіпотенційність системи – властивість системи, що полягає у тому, що кожну систему можна розглядати як підсистему іншої більш крупної системи, а кожен елемент системи в свою чергу є системою.

Емерджентність системи – властивість системи, яка полягає у тому, що складна система має такі властивості, які не притаманні жодному з її елементів.

Ентропія – кількісна міра невизначеності системи.

Ергодичний процес – випадковий процес при якому середнє значення, отримане усередненням однієї реалізації на тривалому проміжку часу співпадає із середнім значенням, отриманим усередненням всіх реалізацій на коротких часових інтервалах.

Завершеність системи – властивість системи, яка виявляється в тому, що вона не допускає приєднання нових елементів без руйнування цієї системи.

Зовнішнє середовище системи – елементи, які не належать системі.

Іманентність системи – властивість системи, яка проявляється в тому, що системоутворююче відношення властиве лише для елементів даної системи.

Інтервал довіри параметрів моделі – інтервали в які із заданою ймовірністю p (рівнем довіри) попадають оцінені нами значення параметрів моделі a_i ($i = \overline{1,m}$).

Кібернетична система – складна динамічна система з управлінням.

Коефіцієнт детермінації – критерій адекватності моделі до статистичних даних, що показує, на скільки відсотків варіація залежної змінної y визначається варіацією незалежної змінної x.

Коефіцієнт конкордації – показник, який визначає ступінь узгодженості думок експертів.

Корелограма – графічне зображенням автокореляційної функції.

Лаг – показник, який характеризує те як вплив зміни пояснюючих змінних на зміну залежної змінної проявляється не миттєво, а з деяким часовим запізненням.

Ланцюг графу – маршрут графу, у якому всі ребра різні.

Простий ланцюг – ланцюг графу, у якого всі вершини різні.

Маршрут графу – скінченна послідовність суміжних ребер.

Математична модель – спрощений абстрактний аналог системи, вираженим за допомогою формул та рівнянь.

Мережа – орієнтований граф, кожна дуга якого має певне числове значення (вагу).

Мінімальність системи – властивість системи, яка проявляється в тому, що система руйнується при вилученні хоча б одного елемента.

Надійність системи – здатність системи зберігати найбільш суттєві властивості на заданому рівні протягом визначеного проміжку часу.

Надійність системи – здатність системи зберігати системоутворюючу властивість при елімінації будь-якої кількості елементів, крім одного.

Орграф (***орієнтований граф***) – граф у якого кожне ребро графа має напрямок.

Пасивні системи – системи, які мають деяке функціональне призначення.

Прогноз – науково обґрунтоване передбачення ймовірності можливих станів системи в майбутньому, шляхів та термінів її розвитку.

Прості системи – системи, що не мають розгалуженої структури, з невеликою кількістю взаємодіючих елементів.

Рівновага системи – здатність системи зберігати свій стан як завгодно довго (як за відсутності, так і за наявності зовнішніх впливів).

Сигнал — це матеріальний засіб перенесення інформації в просторі та часі.

Синергетика –міждисциплінарний науковий напрямок, що вивчає закономірності процесів самоорганізації, еволюції та кооперації.

Синергізм системи – властивість системи, яка характеризується тим, що ефективність спільного функціонування елементів системи є вищою, ніж сумарна ефективність ізольованого функціонування цих же елементів.

Система – множина взаємозв'язаних елементів будь-якої природи, які поєднані за деякими системоутворюючими ознаками та підпорядковані спільній меті.

Системний аналіз – метод пізнання, який використовує розкладання предмету дослідження на складові частини (елементи).

Складні системи – системи з розгалуженою структурою і значною кількістю взаємодіючих елементів.

Стаціонарний процес – випадковий процес при якому із плином часу залишаються незмінними його математичне сподівання та стандартне відхилення**.**

Стійкість системи – здатність системи повертатися до стану рівноваги після виведення її з цього стану.

Структура системи – внутрішня організація системи, яка відображає спосіб взаємодії утворюючих компонентів.

Структура системи – сукупність елементів системи та зв'язків між ними, по яких можуть проходити сигнали.

Траєкторія – лінія, яка з'єднує сусідні точки фазового простору.

Тренд – це стійка систематична зміна процесу протягом тривалого часу.

Управління – цілеспрямоване втручання в перебіг процесів у системі з метою деяких змін або ж підтримання стану системи незмінним.

Фазовий портрет системи – множина усіх траєкторій системи у вигляді сім’ї неперетинних кривих.

Цикл графу – ланцюг графу, у якого початкова і кінцева вершини збігаються.

Цілеспрямовані системи (***активні системи***) – системи, поведінка яких є цілеспрямованою.

ЛІТЕРАТУРА

1. Антонов А. В. Системный анализ : учебник для вузов. Москва : Высшая школа, 2004. 454 с.
2. Богданов А. А. Всеобщая организационная наука (тектология) : в 2-х кн. Москва : Экономика, 1995.
3. Бондарев В. М., Рублинецкий В. И., Качко Е. Г. Основы программирования. Харьков : Фолио, 1998. 368 с.
4. Волкова В. Н., Денисов А. А. Основы теории систем и системного анализа : учебник. 2-е изд. Москва : Юрайт, 2014. 616 с.
5. Волошин О. Ф., Мащенко С. О. Моделі та методи прийняття рішень : навч. посіб. для студ. вищ. навч. закл. Київ : ВПЦ «Київський університет», 2010. 336 с.
6. Гладка О. М., Карпович І. М., Сінчук А. М. Моделі економічної динаміки для фахівців з інформаційних технологій : навч. посіб. Рівне : РДГУ, 2019. 158 с.
7. Грицюк П. М. Аналіз, моделювання та прогнозування динаміки врожайності озимої пшениці в розрізі областей України : монографія. Рівне : НУВГП, 2010. 350 с.
8. Денисов А. А., Колесников Д. Н. Теория больших систем управления. Ленинград : Энергоатомидат, 1992. 288 с.
9. Дивак М. П. Методичний посібник з дисципліни «Системний аналіз». Тернопіль : Тернопільська академія нар. гос-ва, 2004. 136 с.
10. Добровольський В. В. Основи теорії екологічних систем : навч. посіб. Київ : ВД «Професіонал», 2005. 272 с.
11. Задоров В. Б. Системний аналіз об’єктів і процесів: технологічні основи : навч. посіб. Київ : КНУБА, 2003. 276 с.
12. Занг В.-Б. Синергетическая экономика. Время и перемены в нелинейной экономической теории / пер. с англ. Н. Островской и др. Москва : Мир, 1999. 335 с.
13. Згуровський М. З., Панкратова Н. Д. Основи системного аналізу : підручник. Київ : Видавнича група BHV, 2007. 544 с.
14. Калянов Г. Н. CASE структурный системный анализ. Москва : Лори, 1996. 242 с.

15. Катренко А. В. Системний аналіз об'єктів та процесів комп'ютеризації : навч. посіб. Львів : Новий світ–2000, 2007. 424 с.
16. Кузьменко В., Романчук О. На порозі надцивілізації: системний аналіз актуальних проблем сучасності, соціальне прогнозування та футурологія. Львів : Універсум, 2001. 152 с.
17. Кухтенко А. И. Кибернетика и фундаментальные науки. Київ : Наукова думка, 1987. 144 с.
18. Ладанюк А. П. Основи системного аналізу : навч. посіб. Вінниця : Нова книга, 2004. 176 с.
19. Мартинюк П. М., Федорчук Н. А. Теорія системи та математичне моделювання : навч. посіб. Рівне : НУВГП, 2010. 225 с.
20. Месарович М., Такахара И. Общая теория систем: математические основы. Москва : Мир, 1978. 312 с.
21. Основы системного анализа и проектирования АСУ / А. А. Павлов, С. Н. Гриша, В. Н. Томашевский и др. ; под общ. ред. А. А. Павлова. Киев : Выща школа, 1991. 367 с.
22. Перегудов Ф. И., Тарасенко Ф. П. Введение в системный анализ. Москва : Высшая школа, 1992. 367 с.
23. Пригожин И. От существующего к возникающему: Время и сложность в физических науках. Москва : Едиториал УРСС, 2016. 295 с.
24. Рудень В. В., Гутор Т. Г. Методика проведення та оцінки результатів експертних оцінок (на прикладі впровадження системи моніторингу здоров'я населення на рівні первинної медико-санітарної допомоги). *Український медичний часопис*. 2011. № 2. С. 31–34.
25. Саати Т. Принятие решений. Метод анализа иерархий. Москва : Радио и связь, 1993. 278 с.
26. Системный анализ в управлении : учеб. пособ. / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин и др. ; под общ. ред. А. А. Емельянова. Москва : Финансы и статистика, 2002. 368 с.
27. Сорока К. О. Основи теорії систем і системного аналізу : навч. посіб. Харків : ХНАМГ, 2004. 291 с.

28. Спицнадель В. Н. Основы системного анализа. Москва : Бизнес-Пресса, 2000. 326 с.
29. Томашевський В. М. Моделювання систем. Київ : Видавнича група BHV, 2005. 352 с.
30. Шарапов О. Д., Дербенцев В. Д., Семьонов Д. Є. Системний аналіз : навч.-метод. посіб. для самост. вивч. дисц. Київ : КНЕУ, 2003. 154 с.
31. Эшби Р. У. Введение в кібернетику / пер. с англ. Д. Г. Лахути ; под. ред. В. А. Успенского, предисл. А. Н. Колмогорова. 2-е изд. стереотипное. Москва : КомКнига, 2005. 432 с.
32. Юрченко М. Є. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика». Чернігів : ЧНТУ, 2018. 88 с.
33. Hawryszkiewych I. T. Introduction to System Analysis and Design. New York, 1992. 379 p.

Навчальне видання

Грицюк Петро Михайлович
Джоші Олена Іванівна
Гладка Олена Миколаївна

ОСНОВИ ТЕОРІЇ СИСТЕМ І УПРАВЛІННЯ

Навчальний посібник

Друкується в авторській редакції

Технічний редактор Г.Ф. Сімчук

Підписано до друку 26.03.2021 р. Формат 60×84 $^{1}/_{16.}$
Ум.-друк. арк. 15,9. Обл.-вид. арк. 16,6.
Тираж 100 прим. Зам. № 5538.

Видавець і виготовлювач
Національний університет
водного господарства та природокористування
вул. Соборна, 11, м. Рівне, 33028.

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції
РВ № 31 від 26.04.2005 р.