МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Л.Д. Гончарова

МЕТОДИ ОБРОБКИ ТА АНАЛІЗУ ГІДРОМЕТЕОРОЛОГІЧНОЇ ІНФОРМАЦІЇ

Конспект лекцій

Одеса
2017

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський державний екологічний університет

Л.Д. Гончарова

МЕТОДИ ОБРОБКИ ТА АНАЛІЗУ ГІДРОМЕТЕОРОЛОГІЧНОЇ ІНФОРМАЦІЇ

Конспект лекцій

Одеса
2017

ББК 26.23
Г65
УДК. 551.501

Рекомендовано методичною радою Одеського державного екологічного університету Міністерства освіти і науки України як конспект лекцій (протокол № 8 від 25.05.2017р.).

Гончарова Л. Д.
Методи обробки та аналізу гідрометеорологічної інформації: Конспект лекцій. Одеськ. державний екологічний ун-тет. – Одеса; 2017. – 120 с.

Конспект лекцій призначений для студентів, які навчаються за спеціальністю «Науки про Землю» (всіх спеціалізацій). У ньому представлені сучасні методи статистичної обробки результатів випадкових вимірювань та спостережень за атмосферою і гідросферою, опрацьовані за відповідний проміжок часу. Детально розглянуто широке коло питань щодо форм представлення випадкових величин і розрахунків на їх основі точкових та інтервальних статистичних оцінок окремих параметрів генеральної сукупності; дослідження законів розподілу гідрометеорологічних величин. Представлені основні положення щодо перевірки статистичних гіпотез, дослідження однорідності випадкових величин та основ кореляційного зв'язку між двома випадковими величинами.

ББК 26.23
УДК. 551.501

ВСТУП

Загальна сутність екологічних проблем, де б та коли б вони не виникали, полягає в неузгодженості, розладі між тим рівнем перетворення природи, який є історично необхідним для розв'язання проблем соціально-історичного прогресу, які стоять перед людством, і тією широтою та глибиною врахування об'єктивних зв'язків і законів природи, якою людство реально володіє на даному етапі свого розвитку. І в ХХІ столітті людина вперше реально зрозуміла, що вона є жителем планети і може, і повинна думати та діяти в новому – планетарному аспекті.

Виникає запитання: чи треба керувати природою? Так, треба. Але людству дістався в керування дуже складний та норовливий об'єкт. Як стверджував англійський філософ Френсіс Бекон: «Природою можна керувати лише підкоряючись їй».

Один з засновників кібернетики сформулював правило, відповідно якому, керування може бути ефективним лише у тому випадку, коли керуюча система не менш складна, ніж керована. В перекладі на екологічну мову це означає, що ефективно керувати природою можна лише тоді, коли пізнані фізичні закони та взаємозв'язки в оточуючому людину середовищі, а якщо пізнано не все, то керувати можна лише пізнаним. Культура технологічна, що створила людину, потребує збагачення культурою екологічною, яка збереже людину.

У зв'язку з цим практика ставить перед природничими науками задачу пізнання законів, які б дозволили керувати процесами, що відбуваються на Землі в її географічних сферах. Фізичні та хімічні процеси, що виникають і розвиваються в складових кліматичної системи (атмосфера, гідросфера, літосфера, кріосфера, біосфера) вивчаються багатьма науками. Але гідрометеорологічні науки, які входять до складу наук про Землю, відділяються певними особливостями напряму розвитку та методологією досліджень.

Тому в навчальні плани Одеського державного екологічного університету включена спеціальна дисципліна «Методи обробки та аналізу гідрометеорологічної інформації», яка є складовою частиною державного стандарту першого (бакалаврського) рівня вищої освіти. Цей курс є фаховою дисципліною в галузі знань 10 «Природничі науки», спеціальності 103 «Науки про Землю».

Як відомо, гідрометеорологічні величини і явища зумовлюються фізичними процесами різних просторових і часових масштабів – від масштабів, притаманних загальній циркуляції атмосфери чи океану, до масштабів турбулентних вихорів. Чинники, які зумовлюють їх зародження та розвиток, - це ті чи інші прояви законів збереження імпульсу, маси й енергії. Проте їх реалізація у визначений момент часу й у визначеній точці простору часто проявляється як випадкові явища. Тому результати випадкових вимірювань і спостережень, опрацьовані за відповідний проміжок часу, розглядаються як сукупність значень випадкової величини або статистичний ряд. Отже аналіз особливостей змінення тієї чи іншої випадкової величини за часом чи у просторі проводиться як статистичний аналіз випадкових величин.

У природничих науках поняття «статистичний аналіз» означає аналіз масових явищ, що базується на застосуванні методів теорії ймовірності. Пізнання якісних законів розвитку цих явищ неможливе без аналізу їх кількісної сторони. Статистична методологія призначена об'єднати якісну та кількісну сторони складних гідрометеорологічних явищ та процесів і досліджувати сукупність факторів у взаємозв'язку. Вона дозволяє зображувати фізичний процес у цілому, враховуючи тенденції розвитку та різноманітність форм природних об'єктів.

Розуміння основних положень та задач статистичної обробки та аналізу гідрометеорологічної інформації можливе лише на базі знань, здобутих в дисциплінах: «Вища математика (розділ «Теорія ймовірностей та математична статистика»)», «Фізика атмосфери», «Астрономія» та інші.

Стрімка математизація різних сфер людської діяльності, безупинне ускладнення систем та процесів, з якими стикається сучасна людина, приводить до широкого застосування методів математичної статистики і теорії випадкових процесів при вирішенні наукових та практичних задач. Впровадження в практику людської діяльності статистичних методів обробки інформації дозволяє врахувати всю різноманітність зв'язків та факторів, які чинять вплив на процеси, що відбуваються в різних оболонках Землі.

Знайомство з дисципліною «Методи обробки та аналізу гідрометеорологічної інформації» дає можливість вирішувати широке коло питань, які сформовані в розділах: «Поняття про ряди випадкових величин та форми їх представлення», «Статистичні оцінки моментів розподілу випадкових величин», «Інтервальне оцінювання параметрів генеральної сукупності», «Дослідження однорідності випадкових величин», «Апроксимація емпіричних розподілів теоретичними законами», «Дослідження кореляційної залежності між двома випадковими величинами».

Мета вивчення дисципліни полягає у тому, щоб студенти отримали систему теоретичних знань з методів статистичної обробки та аналізу гідрометеорологічної інформації й навичок щодо використання відповідних алгоритмів для розв'язання прикладних задач у науках про Землю.

Після вивчення дисципліни *студент повинен знати*:

– властивості, яким повинні задовольняти сукупності випадкових величин для статистичних досліджень;

– форми подання статистичних сукупностей випадкових величин для розв’язання прикладних задач гідрометеорології;
– методи статистичного оцінювання початкових, центральних, основних моментів розподілу, моди та медіани, їх фізичний сенс;
– алгоритм дослідження закону розподілу випадкової величини; властивості функції розподілу та щільності ймовірності; властивості нормального розподілу, законів Пірсона I, II, III типів, закону Пуассона;
– основи теорії перевірки статистичних гіпотез; методи дослідження однорідності випадкових величин (параметричні та непараметричні критерії);
– побудову довірчих інтервалів для параметрів генеральної сукупності;
– методи побудови рівнянь лінійної регресії для відображення кореляційного зв’язку між двома випадковими величинами.

Студент повинен вміти:

– групувати просту статистичну сукупність та представляти її у табличному та графічному виглядах;
– розраховувати статистичні оцінки початкових, центральних та основних моментів розподілу на основі простих та згрупованих статистичних рядів випадкових величин; розраховувати моду та медіану;
– апроксимувати емпіричний розподіл відомим теоретичним законом (нормальним та законом Пуассона);
– перевіряти статистичні гіпотези щодо однорідності членів статистичного ряду гідрометеорологічних величин, однорідності двох статистичних рядів за допомогою параметричних та непараметричних критеріїв;
– перевіряти статистичну гіпотезу про відповідність емпіричного розподілу визначеному теоретичному закону;
– розраховувати коваріацію та коефіцієнт кореляції;
– розраховувати коефіцієнти лінійного рівняння регресії;
– будувати довірчі інтервали для параметрів генеральної сукупності: математичного сподівання, дисперсії, середнього квадратичного відхилу, коефіцієнта кореляції, коефіцієнтів лінійного рівняння регресії.

Конспект лекцій відповідає робочій програмі дисципліни «Методи обробки та аналізу гідрометеорологічної інформації» і складено на основі використання визначеної кількості першоджерел. Він знайомить студентів з основними особливостями гідрометеорологічної інформації та статистичними методами її обробки з урахуванням специфіки самої інформації, а також сучасних досягнень у галузі статистичної обробки гідрометеорологічних даних.

1. ФОРМА ЗОБРАЖЕННЯ ГІДРОМЕТЕОРОЛОГІЧНОЇ ІНФОРМАЦІЇ

1.1 Особливості гідрометеорологічної інформації

При застосуванні методів теорії ймовірності в гідрометеорологічних дослідженнях використовують статистичні сукупності. Коли кажуть про статистичні сукупності, то мають на увазі дві категорії:

- *генеральна сукупність*;
- *статистичний ряд (вибірка)*.

Термін *«генеральна сукупність»* визначає необмежену кількість незалежних значень випадкової величини, які підпорядковуються одному закону розподілу. Властивості випадкових величин, які представляються генеральною сукупністю, визначаються параметрами цієї випадкової величини. Позначимо їх умовно θ.

Статистичний ряд (вибірка) – обмежена кількість значень випадкової величини, здобутих випадковим чином із генеральної сукупності. Тому статистичні ряди називають вибірками з генеральної сукупності і важливою характеристикою кожної з них є її об'єм (n). Під терміном *«об'єм сукупності випадкової величини»* розуміють кількість членів, що складають цю сукупність. В гідрометеорологічних дослідженнях використовуються ряди як великих ($n>50$), так і малих ($n<50$) об'ємів.

Задача дослідника полягає у тому, щоб за допомогою вибірки розрахувати деякі оцінки параметрів ($\hat{\theta}$), котрі б вірогідно характеризували той чи інший випадковий процес. У зв'язку з цим емпіричні дослідження в науках про Землю мають першорядне значення. На їх основі встановлюються закономірності, притаманні певним характеристикам найбільш мінливих оболонок Землі – атмосфери та гідросфери. Фізичні параметри стану цих двох оболонок кліматичної системи складають гідрометеорологічну інформацію.

Знання комплексу відповідних статистичних алгоритмів і вміння правильно їх використовувати при аналізі цієї інформації допоможе вирішенню актуальних задач утворення, змінення та прогнозування гідрометеорологічних процесів.

Емпіричні дані є критеріями істинності закономірностей, рівнянь гідродинаміки, особливостей метеорологічних чи гідрологічних процесів та інше.

Таким чином, *гідрометеорологічна інформація* має важливі *особливості*, які зумовлюються характером процесів, що спостерігаються в перелічених сферах Землі.

Перша з них полягає в тому, що процеси в океані чи атмосфері мають просторові й часові масштаби, які набагато перевищують можливості окремої людини стосовно збирання та узагальнення інформації про їх стан. Тому дані про процеси в навколишеьому середовищі, що збираються з різних регіонів Землі та за тривалі періоди часу, мають надзвичайну цінність для дослідників.

Друга особливість зумовлюється тим, що в науках про Землю, особливо гідрометеорологічних, є дуже обмежені можливості проведення активного експерименту з природними об'єктами. Отже, аналіз накопичених даних стає головним джерелом досліджень і єдиним засобом перевірки теоретичних висновків та визначених закономірностей.

Особливості об'єктів, що досліджуються, та методів дослідження підкреслюють важливість систем збирання і накопичення гідрометеорологічної інформації та систем забезпечення доступу до неї багатьох користувачів.

Збирання даних про атмосферу та гідросферу здійснюється, по-перше, з метою оперативного доведення інформації до підрозділів гідрометеорологічної служби, які займаються обслуговуванням різних галузей господарства (прогнози погоди, штормові попередження тощо) і, по-друге, для накопичення, з метою узагальнення характеристик про гідрометеорологічний режим та проведення наукових досліджень.

Гідрометеорологічні дані – це кількісні характеристики стану атмосфери та гідросфери. Внаслідок значної мінливості у просторі і за часом фізичних параметрів атмосфери та гідросфери, для спостереження за їх станом з метою вивчення закономірностей процесів, що відбуваються, і, найголовніше, з метою їх прогнозування, необхідні численні вимірювання стану цих середовищ. Відомо, що основним джерелом гідрометеорологічної інформації є результати строкових і спеціальних метеорологічних та гідрологічних спостережень і вимірювань, дані аерологічного та ракетного зондування атмосфери, дані експедиційних досліджень і таке інше.

Значення сукупності гідрометеорологічних величин у певний момент часу визначається станом атмосфери та гідросфери, який обумовлюється дією комплексу фізичних причин. Взагалі кажучи, основні гідрометеорологічні величини є неперервні величини. Це, наприклад, атмосферний тиск, температура і густина повітря, гігрометричні характеристики, швидкість вітру; густина, температура, солоність, швидкість руху води океану тощо. В деяких вимірювальних системах втілюється безперервна регістрація значень тих чи інших фізичних величин. Але в більшості випадків гідрометеорологічні величини вимірюються на світовій мережі метеорологічних чи гідрологічних станцій та постів через деякі проміжки часу, що встановлюються Всесвітньою

Метеорологічною організацією (ВМО) чи особистою програмою досліджень.

Гідрометеорологічні ряди можуть складатися не тільки з величин безпосередньо виміряних. Їх членами можуть бути й величини, отримані в результаті узагальнювання первинних вимірювань чи спостережень.

Треба зауважити, що і у випадку безперервної реєстрації гідрометеорологічної інформації на тих чи інших носіях, перед статистичною обробкою цієї інформації доводиться виконувати її *дискретизацію (квантування).* Цей процес зводиться до складання рядів значень гідрометеорологічної величини у визначені інтервали часу.

Однією з важливих ознак рядів є інтервал дискретності. Як правило, ряди гідрометеорологічних величин є еквідистантними, тобто члени рядів визначаються через будь-який заданий інтервал часу (година, доба, місяць, рік тощо). В деяких випадках при розв'язуванні конкретних задач ряди можуть формуватися з членів, що розташовані на різних відстанях одне від одного.

Важливою властивістю вибірки є характеристика її складових. Такими характеристиками можуть бути безпосередні значення гідрометеорологічних параметрів, кількість днів і випадків з атмосферними явищами, їх тривалість, інтенсивність тощо.

Гідрометеорологічні величини можуть бути скалярними або векторними. В останньому випадку ряд являє собою два або більше (в загальному випадку – ***N***) рядів синхронних скалярних характеристик випадкової величини.

Таким чином, ряди гідрометеорологічних величин складаються з членів, кожен з яких є результатом чи безпосереднього вимірювання або спостереження, чи узагальнювання спостережень за деякий інтервал часу конкретного року.

Отже для гідрометеорологічних досліджень, а також безпосереднього застосування такої інформації в різних галузях господарства, формується велика множина сукупностей гідрометеорологічних величин, які розрізняються однією або декількома ознаками, а саме :

- інтервалом дискретності;
- об'ємом сукупності (вибірки);
- характеристикою значень випадкової величини – членів ряду.

Кожен фізичний параметр атмосфери чи гідросфери залежить один від одного, а також від зовнішніх впливів і випадковим чином змінюється за часом та у просторі, утворюючи випадкові поля або послідовності.

Обробка та аналіз систем випадкових величин проводиться за допомогою спеціально розробленого апарату досліджень, що складає методи математичної статистики. Тому гідрометеорологічна інформація повинна задовольняти вимоги, котрі висуваються до статистичної інформації .

Перш за все, кожен *ряд повинен бути однорідний*. Це означає, що всі члени ряду з визначною ймовірністю повинні належати до однієї генеральної сукупності, тобто підпорядковуватися одному закону розподілу .

У дійсності, в деяких випадках гідрометеорологічні ряди містять члени, які не задовольняють сформульованим вимогам. Їх називають «викидами». «Викиди», як правило, виникають тоді, коли спостерігаються аномальні погодні або кліматичні умови.

Наступною вимогою до рядів гідрометеорологічних величин є *незв'язність* їх *членів*. Це означає, що статистична залежність між ними повинна бути відсутньою. Прийняття чи не прийняття цієї вимоги залежить від характеру задачі, що розв'язується. Якщо йдеться про статистичну оцінку моментів розподілу випадкових величин, то члени вихідних рядів повинні бути незв'язними, оскільки методи статистичного оцінювання параметрів спираються на теореми теорії ймовірностей, які, як правило, ставлять вимогу про незалежність значень випадкової величини.

Зазначену вимогу задовольняють шляхом вибору такого інтервалу дискретності, для якого статистична залежність є незначною (але для цього треба мати апріорну інформацію), або проводити *операцію рандомізації* (від англійського терміну «*random approximation*» – *випадкове наближення*). Для останньої може використовуватися відповідна таблиця випадкових чисел або комп'ютерна програма генерації випадкових чисел.

Інша справа, коли ставиться задача дослідження внутрішньої часової статистичної структури гідрометеорологічних величин. Тоді використовуються вихідні часові ряди визначної дискретності, саме такої, щоб статистична залежність між членами ряду проявлялася в тій чи іншій мірі.

Важливе значення при розрахунках оцінок параметрів має об'єм сукупностей. Статистичний *ряд* повинен *мати представництво*, тобто бути дійсно вибіркою з генеральної сукупності і мати такий об'єм, який дозволяв би провести оцінку параметрів із заданою точністю, тобто отримати вірогідні оцінки. Це можна зробити тільки тоді, коли об'єми вибірок (згідно з так званим законом великих чисел) є досить великими. Але такі сукупності гідрометеорологічних величин не завжди можна сформувати. В гідрометеорології часто виникає ситуація, коли ряди мають малі об'єми. Це зумовлюється терміном, протягом якого організуються спостереження. В такому випадку потрібно проводити оцінку вірогідності отриманого статистичного параметра.

Метеорологічні (або гідрологічні) ряди необхідно подавати у найбільш зручному для аналізу вигляді, в залежності від задачі, що розв'язується. Для статистичних досліджень використовують такі форми їх представлення: у вигляді простого, ранжованого або згрупованого статистичного ряду.

Первинною формою зображення гідрометеорологічної інформації є *простий статистичний ряд*, значення котрого розташовуються в хронологічній послідовності. Такий ряд випадкової величини X об'єму n має вигляд:

$$X : x_1 , x_2 , x_3 , \dots , x_i , \dots , x_n \ , \qquad (1.1)$$

де $x_1 , x_2 , x_3 , \dots , x_i , \dots , x_n$ – конкретні значення випадкової величини *X*, тобто конкретні значення вибірки *X*.

Якщо значення випадкової величини представлені в певній послідовності (в бік збільшення або в бік зменшення їх значень), то такий ряд називають *ранжованим*.

Вихідні дані гідрометеорологічних величин подаються у вигляді *простого* статистичного ряду головним чином у тих випадках, коли задача дослідження полягає у вивченні особливостей їх часової мінливості. Якщо така задача не ставиться, то ряди випадкових величин можуть зображуватися у більш компактній формі – у вигляді *згрупованого ряду*. Особливо це має сенс робити в тому випадку, коли об'єм вибірки є великим, при вирішенні задач з апроксимації емпіричних розподілів теоретичними законами, а також для розв'язання інших прикладних задач гідрометеорології.

1.2 Побудова згрупованого ряду

Нами вже неодноразово підкреслювалося, що інформація про стан процесів в атмосфері чи гідросфері носить випадковий характер та представляється у вигляді випадкових статистичних послідовностей (простих, ранжованих або згрупованих). Задача наук про Землю полягає у тому, щоб з випадкових процесів, які відбуваються в основних оболонках Землі, визначити закономірності, які притаманні цим процесам. Тому для побудови деяких, поки що простих статистичних моделей гідрометеорологічних процесів, використовуються згруповані ряди випадкових величин.

Побудова згрупованого ряду на основі простого статистичного ряду виконується таким чином:

- визначають *область значень* випадкової величини *X* $[x_{min}; x_{max}]$, де x_{min} – найменше, x_{max} – найбільше значення із ряду (1.1);
- всі члени вихідного ряду (1.1) розташовують у новому порядку, а саме в напряму їх збільшення (або зменшення) і така послідовність значень випадкової величини називається *ранжованим* рядом;
- знаходять k – *кількість часткових інтервалів (градацій)*, на які треба поділити область значень випадкової величини *X* $[x_{min}; x_{max}]$. Для цього може використовуватися формула:

$$k = 5\lg n \ , \tag{1.2}$$

де n – *об'єм ряду;* за k завжди беруть *ціле число*.

Для зазначеної мети можна також використовувати формулу:

$$k = 1 + 3,222\lg n \ ; \tag{1.3}$$

- далі знаходять *довжину часткового інтервалу* c за формулою:

$$c = \frac{x_{max} - x_{min}}{k} \ . \tag{1.4}$$

Розрахунки довжини градації проводять з тією точністю, з якою представлені значення випадкової величини;

- потім визначають *значення* випадкової величини X *на межах часткових інтервалів*. Для i-го часткового інтервалу, очевидно, значення величини X *на лівій межі* є $\left[x_{min} + (i-1)c\right]$, а *на правій* – $\left[x_{min} + ic\right]$. Умовно будемо позначати *ліву* межу часткового інтервалу як x_{i-1}, а *праву* – x_{i+1} $\left(i = \overline{1,k}\right)$.

Примітка: кінець попередньої і початок наступної градацій будуть повторюватися. Тому треба визначити закриті й відкриті межі градацій, тобто встановити, яку з величин враховувати в даній градації, щоб виключити повторення одних і тих же значень випадкової величини X *в різних градаціях;*

- підраховують *кількість членів ряду*, що потрапляють до кожного i-го часткового інтервалу – m_i $\left(i = 1, 2, \ldots, k\right)$. Величини m_i називають *інтервальними емпіричними частотами*. Зрозуміло, що сума частот по всіх часткових інтервалах (накопичені частоти) дорівнює об'єму вибірки *X*, тобто

$$\sum_{i=1}^{k} m_i = n \, . \tag{1.5}$$

Накопичені частоти відносяться не до середин градацій, а до меж інтервалів;

- розраховують *інтервальні частості* $\hat{p}_i$ (*відносні інтервальні частоти*) за формулою:

$$\hat{p}_i = \frac{m_i}{n} \, . \tag{1.6}$$

Інтервальна частість – це ймовірність попадання значення випадкової величини в той чи інший інтервал.

Відносна частота часто виражається у відсотках:

$$\hat{p}_i = \frac{m_i}{n} \cdot 100\% \ . \tag{1.7}$$

Очевидно, сума частостей по всіх k градаціях (накопичені ймовірності) дорівнює одиниці або 100%:

$$\sum_{i=1}^{k} \hat{p}_i = 1 \quad \text{або} \quad \sum_{i=1}^{k} \hat{p}_i = 100\% \, ; \tag{1.8}$$

- знаходять $\tilde{x}_i$ – значення випадкової величини X на *середині* кожного *часткового інтервалу* за формулою:

$$\tilde{x}_i = \frac{x_{i-1} + x_{i+1}}{2} \quad \left(i = \overline{1,k}\right). \tag{1.9}$$

Після останнього етапу всі ознаки згрупованого ряду є й можна дати його визначення.

Згрупованим називають такий *ранжований статистичний ряд,* який можна аналітично представити сукупністю значень випадкової величини X на серединах часткових інтервалів $\tilde{x}_i$ і відповідних інтервальних частот m_i :

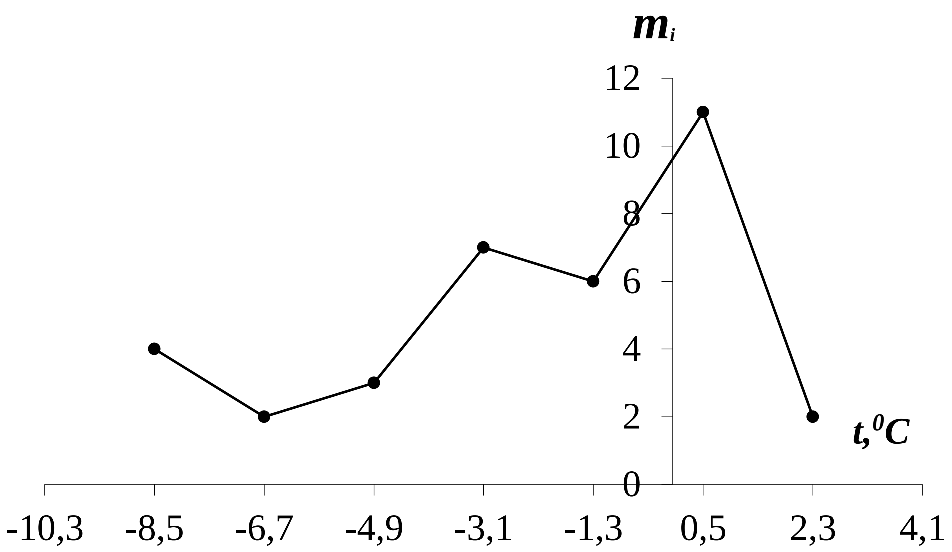
$$X: \begin{cases} \tilde{x}_1, \tilde{x}_2 , \tilde{x}_3, ..., \tilde{x}_i ,..., \tilde{x}_k \\ m_1, m_2 , m_3 ,..., m_i ,..., m_k \end{cases} \quad \left(i = \overline{1,k}\right). \tag{1.10}$$

Згрупований статистичний ряд випадкової величини X аналітично можна записати і таким чином:

$$X: \begin{cases} \tilde{x}_1, \tilde{x}_2, \tilde{x}_3,..., \tilde{x}_i,..., \tilde{x}_k \\ \hat{p}_1, \hat{p}_2, \hat{p}_3,..., \hat{p}_i,..., \hat{p}_k \end{cases} \quad \left(i = \overline{1,k}\right), \tag{1.11}$$

де $\tilde{x}_i$ – середина, а $\hat{p}_i$ – інтервальна частість i -ої градації.

Згрупований ряд (по сенсу) є емпіричним розподілом інтервальних імовірностей випадкової величини і може зображуватися графічно за допомогою *гістограми* чи *полігону*. *Гістограма* – це система прямокутників, основи яких дорівнюють довжині часткового інтервалу, а висоти – відповідним інтервальним частотам (або частостям). Якщо точки з координатами $\left(\tilde{x}_i; m_i\right)$ або $\left(\tilde{x}_i; \hat{p}_i\right)$ з’єднати відрізками прямої, то отриману таким чином діаграму називають *полігоном*. Приклади цих діаграм представлено на рис. 1.1.



а)

б)

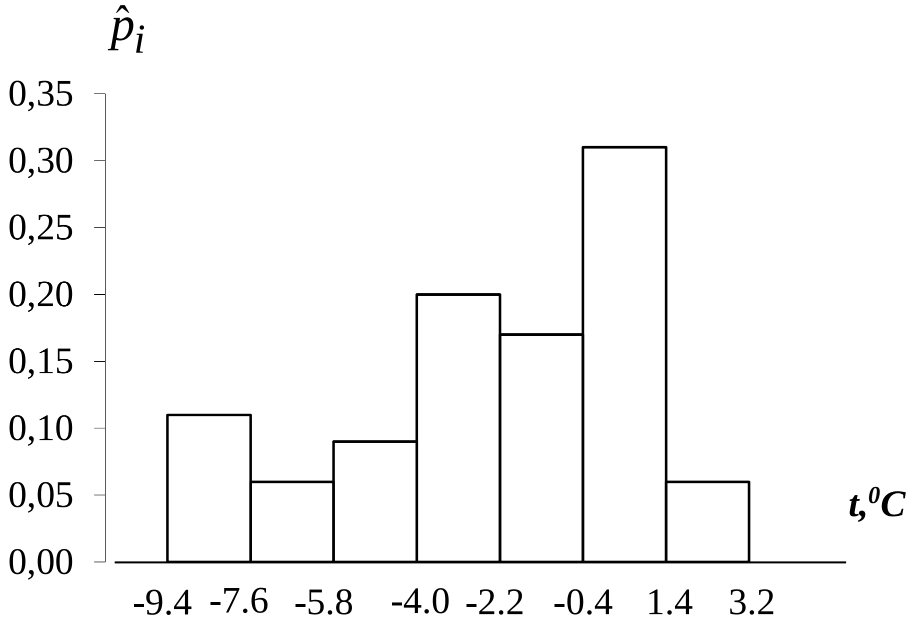


Рис. 1.1 – Полігон (а) та гістограма (б) розподілу середньомісячної температури повітря (ст. Одеса, січень)

Запитання для самоперевірки

1. У чому відмінність між генеральною сукупністю та статистичним рядом (вибіркою)?
2. Перелічити форми зображення статистичних рядів (вибірок).
3. Яким вимогам повинна відповідати гідрометеорологічна інформація?
4. Який ряд значень випадкової величини називають «простим статистичним рядом»? Яким виразом він може бути представлений?
5. Який ряд значень випадкової величини називають «ранжованим статистичним рядом»?
6. Який ряд значень випадкової величини називають «згрупованим статистичним рядом»?
7. Від яких величин залежить кількість градацій у згрупованій сукупності?
8. Від яких величин залежить довжина часткового інтервалу?
9. Дати визначення «інтервальної емпіричної частоти». Як за допомогою цієї величини знайти об'єм вибірки?
10. Дати визначення «об'єму вибірки».
11. Дати визначення «інтервальної частості». Чому дорівнює сума частостей по всіх градаціях?
12. Як графічно можна представити згруповані ряди?
13. Яка діаграма розподілу випадкової величини називається «гістограмою»?
14. Яка діаграма розподілу випадкової величини називається «полігоном»?
15. Які важливі особливості має гідрометеорологічна інформація?
16. Якими ознаками характеризується сукупність гідрометеорологічної величини?

2. ТОЧКОВІ СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ

Як відомо [2, 9, 10], важливі особливості просторової та часової мінливості *випадкових величин*, якими є гідрометеорологічні величини, вивчаються на основі їх статистичних сукупностей.

Нагадаємо, що термін *«генеральна сукупність»* визначає нескінченну множину незалежних значень випадкової величини, які підпорядковуються одному й тому ж закону розподілу. *Властивості випадкових величин, які представляються генеральною сукупністю, визначаються параметрами* цієї випадкової величини. Позначили їх умовно $\boldsymbol{\theta}$. Параметрами генеральної сукупності є математичне сподівання m_x, дисперсія $\boldsymbol{\sigma}_x^2$, середній квадратичний відхил $\boldsymbol{\sigma}_x$ та інші.

Статистичний ряд (вибірка) – обмежена кількість значень випадкової величини, добута випадковим чином із генеральної сукупності. Тому статистичні ряди називають вибірками з генеральної сукупності. Вибірки випадкові та кількість їх безмежна. *Статистичні ряди* гідрометеорологічних величин називають *випадковими величинами* і позначають великими літерами латинської абетки (X, Y, Z …), а їх конкретні значення – відповідно малими літерами (x_i , y_i , $z_i, \ldots$).

Важливою ознакою ряду є його *об'єм*. Під терміном *«об'єм»* сукупності значень випадкової величини розуміють кількість членів, які складають цю сукупність. В гідрометеорологічних дослідженнях доводиться використовувати ряди як великих, так і малих об'ємів, оскільки вони формуються шляхом безпосередніх спостережень або вимірювань, або шляхом деяких первинних узагальнень.

Вибірки характеризуються статистичними оцінками параметрів.

Значення параметра генеральної сукупності, яке розраховують на основі вибірки, називають статистичною оцінкою ($\hat{\boldsymbol{\theta}}$) параметра $\boldsymbol{\theta}$ і позначають символом "^".

З теорії ймовірності відомо, що основні властивості випадкових величин характеризуються *початковими* (ν_l), *центральними* (μ_l) та *основними* (r_l) моментами розподілу різних порядків (*l*). В гідрометеорологічних дослідженнях, як правило, використовуються перелічені моменти перших чотирьох порядків ($l = \overline{1,4}$). Початкові, центральні та основні моменти розподілу є параметрами генеральних сукупностей випадкових величин. На основі статистичної сукупності (вибірки) можна знайти *статистичні оцінки* цих параметрів.

Будемо у подальшому позначати статистичні оцінки моментів розподілу l-го порядку таким чином: початкового – $\hat{\nu}_l$, центрального – $\hat{\mu}_l$, основного – $\hat{r}_l$. Вони можуть розраховуватися як по простих, так і по згрупованих статистичних сукупностях різних гідрометеорологічних величин (наприклад, температури повітря, місячної кількості опадів, атмосферного тиску, швидкості вітру і т.п.). За статистичною методологією отримані по вибірках оцінки параметрів повинні задовольняти певним умовам.

2.1 Властивості статистичних оцінок параметрів

2.1.1 Умови, які повинні задовольняти статистичні оцінки параметрів

Нехай з генеральної сукупності здобуто випадковим чином N вибірок. Тоді будемо мати N оцінок параметра. Якими повинні бути оцінки $\hat{\theta}_N$, щоб достатньо вірогідно характеризувати параметр θ? Вони повинні задовольняти *три умови: незсуненості, ефективності* та *умотивованості.*

Оцінка параметра називається *незсуненою*, якщо її математичне сподівання дорівнює параметру, який оцінюється, тобто

$$M\left[\hat{\theta}_N\right]=\theta. \tag{2.1}$$

Якщо ця рівність не виконується, то оцінка $\hat{\theta}_N$ може або завищувати значення θ (тобто $M[\hat{\theta}_N]>\theta$), або занижувати його ($M\left[\hat{\theta}_N\right]<\theta$). В обох випадках це призводить до систематичних (одного знаку) похибок в оцінці параметра θ. Отже вимога незсуненості гарантує відсутність систематичних похибок при оцінках параметрів.

Оскільки $\hat{\theta}_N$ – випадкова величина, значення якої змінюється від вибірки до вибірки (нагадаємо, що вибірки вибираються з генеральної сукупності випадковим чином), то міра її розсіювання відносно математичного сподівання параметра θ характеризується дисперсією $Д[\hat{\theta}_N]$. Нехай $\hat{\theta}_N$ і $\hat{T}_N$ – дві незсунені оцінки параметра θ, тобто $M[\hat{\theta}_N]=\theta$; $M[\hat{T}_N]=\theta$. Із двох оцінок $\hat{\theta}_N$ і $\hat{T}_N$ слід віддати перевагу тій, яка має менше розсіювання навколо оцінюваного параметра, тому що в цьому випадку значення оцінки, що знайдене з конкретної вибірки, нвйменше відхиляється від істинного значення параметра θ. Якщо $Д[\hat{\theta}_N] < Д[\hat{T}_N]$, то за оцінку треба брати $\hat{\theta}_N$. Незсунена оцінка, котра має найменшу дисперсію серед усіх можливих незсунених оцінок параметра, розрахованих по вибірках одного й того ж об'єму, називається *ефективною оцінкою*. Іншими словами, оцінка $\hat{\theta}_N$ більш ефективна, ніж $\hat{T}_N$, якщо $M[(\hat{\theta}_N - \theta)^2] < M[(\hat{T}_N - \theta)^2]$.

Розглянемо нижню межу $\inf M[(\hat{\theta}_N - \theta)^2]$ множини величин $M[(\hat{\theta} - \theta)^2]$ по всіх можливих $\hat{\theta}_N$. Оцінка $\hat{\theta}_N$, для якої досягається нижня межа, називається ефективною. Оцінка $\hat{\theta}$, для якої міра розкиду $M[(\hat{\theta} - \theta)^2]$ при $N \to \infty$ всюди поводиться так, як і для ефективної оцінки, тобто

$$\lim_{N \to \infty} \frac{M[(\hat{\theta}_N - \theta)^2]}{M[(\hat{\theta} - \theta)^2]} = 1 \qquad (2.2)$$

називається *асимптотичною ефективною оцінкою*.

Оцінка $\hat{\theta}_N$ параметра θ називається *умотивованою*, якщо вона по ймовірності збігається до параметра θ:

$$\lim_{n \to \infty} (P\left|\hat{\theta} - \theta\right| < \varepsilon) = 1. \qquad (2.3)$$

Рівність (2.3) позначає, що чим більший об'єм вибірки (n), тим більша ймовірність того, що похибка оцінки не перевищує скільки завгодно малого додатного числа. Отже, у випадку використання умотивованих оцінок виправдовується збільшення числа одиниць вибірки, оскільки при цьому все менше ймовірною стає можливість значної помилки в оцінці невідомого параметра.

Коли $Д[\hat{\theta}_N] \to 0$ при $n \to \infty$, то незсунена оцінка є й ефективною, й умотивованою.

2.1.2 Середнє значення випадкової величини як незсунена, ефективна та умотивована оцінка математичного сподівання

Покажемо, що *середнє значення випадкової величини є незсуненою, ефективною та умотивованою оцінкою математичного сподівання*. Для цього знайдемо математичне сподівання середнього значення випадкової величини X:

$$M[\bar{x}] = M\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]. \qquad (2.4)$$

За відомими властивостями математичного сподівання маємо:

$$M[\overline{x}] = \frac{1}{n}\sum_{i=1}^{n} M\left[x_i\right] = \frac{1}{n}\sum_{i=1}^{n} m_x = m_x . \qquad (2.5)$$

Рівність (2.5) і є визначення незсуненості оцінки математичного сподівання m_x, тобто $\overline{x}$.

Знайдемо тепер дисперсію середнього значення

$$Д[\overline{x}] = Д\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]. \qquad (2.6)$$

У відповідності до властивостей дисперсії можна записати

$$Д[\overline{x}] = \frac{1}{n^2}\sum_{i=1}^{n} Д[x_i] = \frac{1}{n^2}\sum_{i=1}^{n} {\sigma_x}^2 = \frac{{\sigma_x}^2}{n} . \qquad (2.7)$$

Оскільки при $n \to \infty$, $Д[\overline{x}] \to 0$, то середнє значення є ефективною та умотивованою оцінкою математичного сподівання.

Очевидно

$$\sqrt{Д[\overline{x}]} = \sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}} . \qquad (2.8)$$

Формула (2.8) характеризує середній квадратичний відхил середнього значення.

2.1.3 Частість як незсунена, ефективна та умотивована оцінка ймовірності

Нехай проводяться експерименти з деяким явищем А, яке може з'явитися, а може не з'явитися. Знайдемо z_ν – кількість появ події A в одній ν-ій спробі. Очевидно, z_ν, якщо подія А відбулася з імовірністю p і $z_\nu = 0$, якщо подія А не відбулася з імовірністю $q = 1 - p$.

Знайдемо математичне сподівання кількості появ події А:

$$M[z_\nu] = 1 \cdot p + 0 \cdot q = p , \qquad (2.9)$$

а також дисперсію

$$Д[z_\nu] = (1-p)^2 p + (0-p)^2 \cdot q = q^2 p + p^2 q = pq(p+q) = pq . \qquad (2.10)$$

Частоту m можна виразити через кількість появ події А таким чином:

$$m = \sum_{\nu=1}^{n} z_\nu , \qquad (2.11)$$

де n - загальна кількість незалежних експериментів.

Тоді частість $\hat{p}$ дорівнює:

$$\hat{p} = \frac{\sum\limits_{\nu=1}^{n} z_\nu}{n} . \tag{2.12}$$

Знайдемо математичне сподівання частості:

$$M[\hat{p}] = M\left[\frac{\sum\limits_{\nu=1}^{n} z_\nu}{n}\right] = \frac{1}{n}\sum_{\nu=1}^{n} M[z_\nu] . \tag{2.13}$$

Урахування співвідношення (2.9) приводить до результату:

$$M[\hat{p}] = \frac{1}{n}\sum_{\nu=1}^{n} p = p . \tag{2.14}$$

Рівність (2.14) свідчить про те, що частість є незсуненою оцінкою ймовірності.

Розрахуємо тепер дисперсію частості. Будемо мати:

$$Д[\hat{p}] = Д\left[\frac{1}{n}\sum_{\nu=1}^{n} z_\nu\right] = \frac{1}{n^2}\sum_{\nu=1}^{n} Д\left[z_\nu\right]. \tag{2.15}$$

Підставимо (2.15) в результат (2.10). Маємо:

$$Д[\hat{p}] == \frac{1}{n^2}\sum_{i=1}^{n} pq = \frac{pq}{n} . \tag{2.16}$$

Очевидно при $n \to \infty$, $Д\left[\hat{p}\right] \to 0$. Отже частість є незсуненою, ефективною та умотивованою оцінкою ймовірності.

2.2 Початкові моменти розподілу та їх статистичні оцінки

З теорії ймовірностей відомо [8], що властивості випадкових величин можуть характеризуватися початковими (ν), центральними (μ) та основними (r) моментами різних порядків (l) .

В гідрометеорологічних дослідженнях, як правило, використовуються перелічені моменти перших чотирьох порядків, які, як буде показано пізніше, відбивають фізичні властивості випадкових процесів.

Початковий момент l-го порядку для безперервної випадкової величини X визначається таким чином:

$$\nu_l = \int_{-\infty}^{\infty} x^l f(x)dx \ , \tag{2.17}$$

де $f(x)$ – щільність імовірності випадкової величини X .

На основі цього визначення отримаємо метод, за допомогою якого можна знайти статистичну оцінку l-го початкового моменту на основі вибірки випадкової величини, яка може бути задана згрупованим рядом вигляду (1.10) або (1.11).

Як випливає з формули (2.17), випадкова величина X визначена на інтервалі $(-\infty,\infty)$. Інтервал же значень випадкової величини, що визначається вибіркою $X: x_1, x_2, x_3, ..., x_n$, є обмеженим $[x_{min}; x_{max}]$. Тому будемо моделювати випадкову величину X таким чином:

$$X = \begin{cases} 0, & \text{за умови} \quad x < x_{min}, \\ x, & \text{за умови} \quad x_{min} \le x \le x_{max}, \\ 0, & \text{за умови} \quad x > x_{max}. \end{cases} \tag{2.18}$$

З урахуванням умов (2.18), запишемо інтеграл (2.17) у вигляді :

$$\nu_l = \int_{-\infty}^{x_{min}} x^l f(x)dx + \int_{x_{min}}^{x_{max}} x^l f(x)dx + \int_{x_{max}}^{\infty} x^l f(x)dx \ . \tag{2.19}$$

Перший та третій інтеграли у формулі (2.19) дорівнюють нулю.
Отже,

$$\nu_l = \int_{x_{min}}^{x_{max}} x^l f(x)dx \, . \tag{2.20}$$

При побудові згрупованого ряду інтервал $[x_{min}; x_{max}]$ роздрібнювався на k часткових інтервалів довжиною c. Враховуючи це, розіб'ємо інтеграл (2.20) на k інтегралів

$$\nu_l = \sum_{i=1}^{k} \int_{x_{min}+(i-1)c}^{x_{min}+ic} x^l f(x)dx \ . \tag{2.21}$$

Розглянемо інтервал $[x_{min} + (i-1)c; \ x_{min} + ic]$ (рис. 2.1). Очевидно, при апроксимації криволінійної трапеції $AB'C'Д$ прямокутником $ABCД$ середнє значення випадкової величини X в цьому інтервалі є її значення x_i на середині i-го часткового інтервалу ($\tilde{x}_i$), а середнє значення щільності ймовірності є її значення в точці $\tilde{x}_i$ – $f(\tilde{x}_i)$.

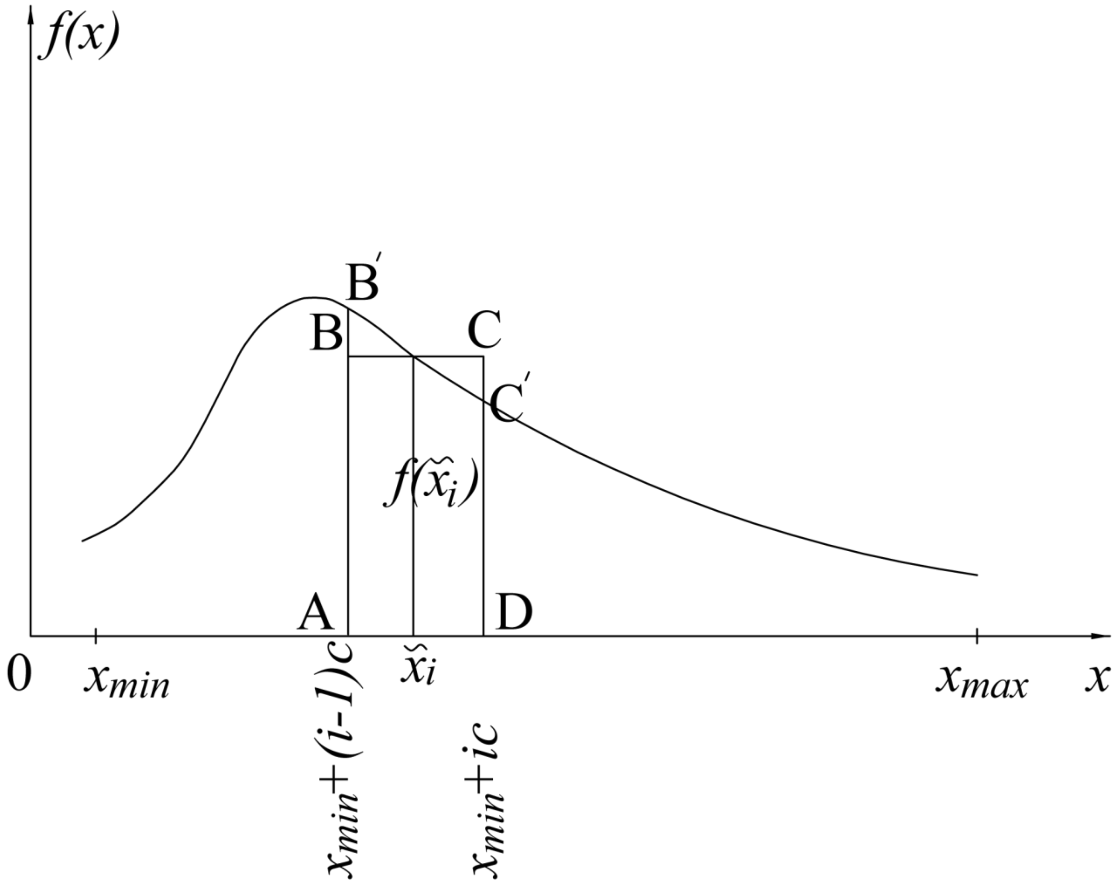


Рис. 2.1 – До виведення формул статистичних оцінок моментів розподілу випадкової величини ***X*** [9]

З урахуванням цього, застосуємо до інтеграла в формулі (2.21) теорему про середнє. Будемо мати:

$$v_l \approx \sum_{i=1}^{k} \tilde{x}_i^{\,l} f(x_i) \int_{x_{min}+(i-1)c}^{x_{min}+ic} dx = \hat{v}_l \qquad (2.22)$$

або

$$\hat{v}_l \approx \sum_{i=1}^{k} \tilde{x}_i^{\,l} f(x_i) c\,. \qquad (2.23)$$

Добуток

$$f(\tilde{x}_i) c = \hat{p}_i \qquad (2.24)$$

є площа прямокутника $АВСД$, яка приблизно дорівнює площі криволінійної трапеції $АВ^{'}С^{'}Д$, що відображає інтервальну ймовірність $\hat{p}_i$ випадкової величини X. Тому можна вважати, що величина $\hat{p}_i$, яка визначається формулою (2.24), є оцінкою цієї ймовірності, тобто інтервальною частістю.

Таким чином, якщо для знаходження використовуються згруповані ряди вигляду (1.10) або (1.11), статистична оцінка l-го початкового моменту дорівнює

$$\hat{\nu}_l = \sum_{i=1}^{k} \tilde{x}_i^{\,l} \hat{p}_i \tag{2.25}$$

або, оскільки $\hat{p}_i = \dfrac{m_i}{n}$,

$$\hat{\nu}_l = \frac{1}{n} \sum_{i=1}^{k} \tilde{x}_i^{\,l} m_i , \tag{2.26}$$

де n – об’єм вибірки;

k – кількість часткових інтервалів;

$\tilde{x}_i, m_i, \hat{p}_i$ – середина, інтервальна емпірична частота та інтервальна частість i-ї градації відповідно.

Із теорії ймовірності відомо, що початковий момент першого порядку

$$\hat{\nu}_1 = \int_{-\infty}^{\infty} x f(x) dx = m_x \tag{2.27}$$

є математичним сподіванням випадкової величини X.

Очевидно, *оцінка першого початкового моменту розподілу* $(\hat{\nu}_1)$ *є оцінкою математичного сподівання* $(\hat{m}_x)$ і дорівнює *середньому значенню* $(\overline{x})$ випадкової величини X:

$$\hat{\nu}_1 = \hat{m}_x = \overline{x} = \frac{1}{n} \sum_{i=1}^{k} \tilde{x}_i m_i \tag{2.28}$$

або

$$\hat{\nu}_1 = \hat{m}_x = \overline{x} = \sum_{i=1}^{k} \tilde{x}_i \hat{p}_i . \tag{2.29}$$

Статистичні оцінки початкових моментів розподілу *другого*, *третього* та *четвертого* порядків $\left(\hat{\nu}_2, \hat{\nu}_3 \text{ та } \hat{\nu}_4\right)$ розраховуються за формулами (2.25) або (2.26) за умов, що в цих формулах показник степеня l набуває відповідних значень 2, 3 і 4:

$$\hat{\nu}_2 = \frac{1}{n} \sum_{i=1}^{k} \tilde{x}_i^2 m_i = \overline{x^2}, \tag{2.30}$$

$$\hat{v}_3 = \frac{1}{n}\sum_{i=1}^{k} \tilde{x}_i^3 m_i = \overline{x^3}, \quad (2.31)$$

$$\hat{v}_4 = \frac{1}{n}\sum_{i=1}^{k} \tilde{x}_i^4 m_i = \overline{x^4}\,. \quad (2.32)$$

Якщо статистична оцінка l-го *початкового* моменту розподілу обчислюється на основі *простих статистичних сукупностей* вигляду (1.1), то використовується формула:

$$\hat{v}_l = \frac{1}{n}\sum_{i=1}^{n} x_i^l\,. \quad (2.33)$$

Очевидно, *середнє значення* для таких сукупностей визначимо за допомогою рівняння:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{v}_1 = \hat{m}_x, \quad (2.34)$$

а оцінки початкових моментів розподілу 2-го, 3-го та 4-го порядків – за формулами:

$$\hat{v}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \overline{x^2}, \quad (2.35)$$

$$\hat{v}_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3 = \overline{x^3}, \quad (2.36)$$

$$\hat{v}_4 = \frac{1}{n}\sum_{i=1}^{n} x_i^4 = \overline{x^4}\,. \quad (2.37)$$

У рівняннях (2.33) - (2.37) n – об’єм вибірки; x_i – кожне конкретне значення випадкової величини X із ряду (1.1).

2.3 Центральні моменти розподілу та їх статистичні оцінки

Центральним моментом розподілу l–го порядку випадкової величини *X* називається інтеграл вигляду:

$$\mu_l = \int_{-\infty}^{\infty} (x - m_x)^l f(x)dx\,. \quad (2.38)$$

Центральні моменти розподілу оцінюються, починаючи з другого моменту (l=2), тому, що *перший центральний момент завжди дорівнює нулю, як і його оцінка.* Центральний момент другого порядку має сенс дисперсії випадкової величини:

$$\mu_2 = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx = \sigma_x^2 . \qquad (2.39)$$

Аналогічним чином (як це показано у підрозділі 2.2 для статистичних оцінок початкових моментів розподілу) можна отримати с*татистичну оцінку центрального моменту розподілу* l*-го порядку* для випадкової величини *X,* якщо ряди є *згрупованими*:

$$\hat{\mu}_l = \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^l \hat{p}_i \qquad (2.40)$$

або при використанні інтервальних емпіричних частот m_i за формулою:

$$\hat{\mu}_l = \frac{1}{n} \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^l m_i . \qquad (2.41)$$

Для розрахування статистичної оцінки *центрального моменту розподілу другого* порядку на основі згрупованого ряду використовуються формули:

$$\hat{\mu}_2 = \hat{\sigma}_x^2 = \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^2 \hat{p}_i , \qquad (2.42)$$

$$\hat{\mu}_2 = \hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^2 m_i . \qquad (2.43)$$

Аналогічним чином на основі формул (2.40) або (2.41) визначають статистичні оцінки *третього* (за умови l =3) і *четвертого* (за умови l =4) моментів розподілу випадкової величини X :

$$\hat{\mu}_3 = \frac{1}{n} \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^3 m_i , \qquad (2.44)$$

$$\hat{\mu}_4 = \frac{1}{n} \sum_{i=1}^{k} \left(\tilde{x}_i - \overline{x} \right)^4 m_i . \qquad (2.45)$$

Статистична оцінка другого центрального моменту розподілу, що розраховується за формулою (2.42) або (2.43), є *зсуненою оцінкою* дисперсії.

Незсунена, *ефективна* та *умотивована* оцінка дисперсії випадкової величини *X* (це вимоги, яким повинні задовольняти статистичні оцінки параметрів) пов'язана з центральнім моментом розподілу 2-го порядку формулою (2.46). Її позначають S_x^2:

$$S_x^2 = \frac{n}{n-1} \hat{\mu}_2 , \qquad (2.46)$$

де $\left(\dfrac{n}{n-1}\right)$ – коефіцієнт Бесселя.

Щоб розрахувати незсунену, ефективну та умотивовану оцінку дисперсії випадкової величини *X* за умови згрупованих рядів використовують рівняння (2.47):

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{k}(\tilde{x}_i - \overline{x})^2 m_i \,. \tag{2.47}$$

Очевидно, статистична оцінка *середнього квадратичного відхилу* цієї величини (S_x) визначається як:

$$S_x = \sqrt{S_x^2}\,. \tag{2.48}$$

Статистичну оцінку *центрального моменту розподілу* l*-го* порядку у випадку *простих статистичних сукупностей* випадкової величини *X* отримаємо за допомогою формули:

$$\hat{\mu}_l = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^l . \tag{2.49}$$

Тоді статистичні оцінки *центральних моментів другого, третього та четвертого* порядків розраховуються за формулами (2.50) - (2.52):

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \hat{\sigma}_x^2 \,, \tag{2.50}$$

$$\hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3 , \tag{2.51}$$

$$\hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4 , \tag{2.52}$$

де – n – об’єм ряду;

$x_i\left(i = \overline{1,n}\right)$ – конкретне значення випадкової величини *X*;

$\overline{x}$ – середнє значення вибірки *X*.

Незсунена, ефективна та умотивована оцінка дисперсії випадкової величини *X*, що задана простим рядом, визначається за формулою:

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 . \tag{2.53}$$

Центральні моменти розподілу 2-го, 3-го та 4-го порядків можна розрахувати і за формулами їх зв'язку з початковими моментами:

$$\hat{\mu}_2 = \hat{\nu}_2 - \hat{\nu}_1^2 , \tag{2.54}$$

$$\hat{\mu}_3 = \hat{\nu}_3 - 3\hat{\nu}_2\hat{\nu}_1 + 2\hat{\nu}_1^3, \tag{2.55}$$

$$\hat{\mu}_4 = \hat{\nu}_4 - 4\,\hat{\nu}_1\hat{\nu}_3 + 6\,\hat{\nu}_1^2\hat{\nu}_2 - 3\hat{\nu}_1^4 . \tag{2.56}$$

2.4 Модальне значення та медіана випадкової величини

Інколи при статистичних дослідженнях гідрометеорологічних рядів необхідно визначити їх модальне значення Mo та медіану Me. Особливо ці статистичні оцінки є важливими в задачах добору до емпіричного закону теоретичного розподілу.

Мода – значення випадкової величини, що зустрічається найчастіше, тобто має максимальну ймовірність для дискретної величини або максимум функції щільності ймовірності в даній точці для неперервної випадкової величини. Одна й та ж випадкова величина може мати одну або декілька мод. Однак можливо, що випадкова величина і не матиме моди, якщо її значення мають однакову ймовірність (рівномірний розподіл) або, якщо ймовірність при зростанні значення випадкової величини безперервно зростає або зменшується.

Якщо ряд *згрупований* і *ранжування* проводилося *в бік зростання* значень випадкової величини, то моду визначають за такою формулою:

$$M_o = x_0 + c\frac{\left(m_i - m_{i-1}\right)}{\left(2m_i - m_{i-1} - m_{i+1}\right)}, \tag{2.57}$$

де x_0, c, m_i – відповідно початок, довжина та емпірична частота модального інтервалу;

$m_{i-1},\ m_{i+1}$ – частоти попереднього і наступного за модальним часткових інтервалів.

Якщо випадкова величина має одну моду то розподіл називають одномодальним; дві моди – двомодальним (бімодальним) розподілом.

Медіана – значення випадкової величини, яке розділяє область існування цієї величини на дві рівні по ймовірності частини, для яких виконується співвідношення:

$$P\left(X > Me\right) = P\left(X < Me\right), \tag{2.58}$$

тобто ймовірність того, що значення випадкової величини X більше медіани, дорівнює ймовірності того, що значення випадкової величини X менше медіани.

Медіану розраховують на основі згрупованого ряду за такою формулою:

$$Me = x_e + \frac{c\left(\frac{n}{2} - m^*\right)}{m_e}, \tag{2.59}$$

де

n – об'єм вибірки (за умови непарного значення об'єму береться $n+1$);

m^* – накопичена частота до медіанного інтервалу;

x_e, c, m_e – відповідно початок, довжина та частота медіанного інтервалу.

2.5 Основні моменти розподілу та їх статистичні оцінки

Основний (нормований) момент розподілу l-го порядку є часткою від ділення l-го центрального моменту на середній квадратичний відхил в l-му степені:

$$r_l = \frac{\mu_l}{\sigma_x^l}. \tag{2.60}$$

Статистична оцінка основного моменту l-го порядку для випадкової величини X розраховується за формулою:

$$\hat{r}_l = \frac{\hat{\mu}_l}{S_x^l}. \tag{2.61}$$

Очевидно, що $r_1 = 0$, а $r_2 = 1$.

Тому обчислюються тільки статистичні оцінки *третього* (за умови l =3):

$$\hat{r}_3 = \frac{\hat{\mu}_3}{S_x^3} \tag{2.62}$$

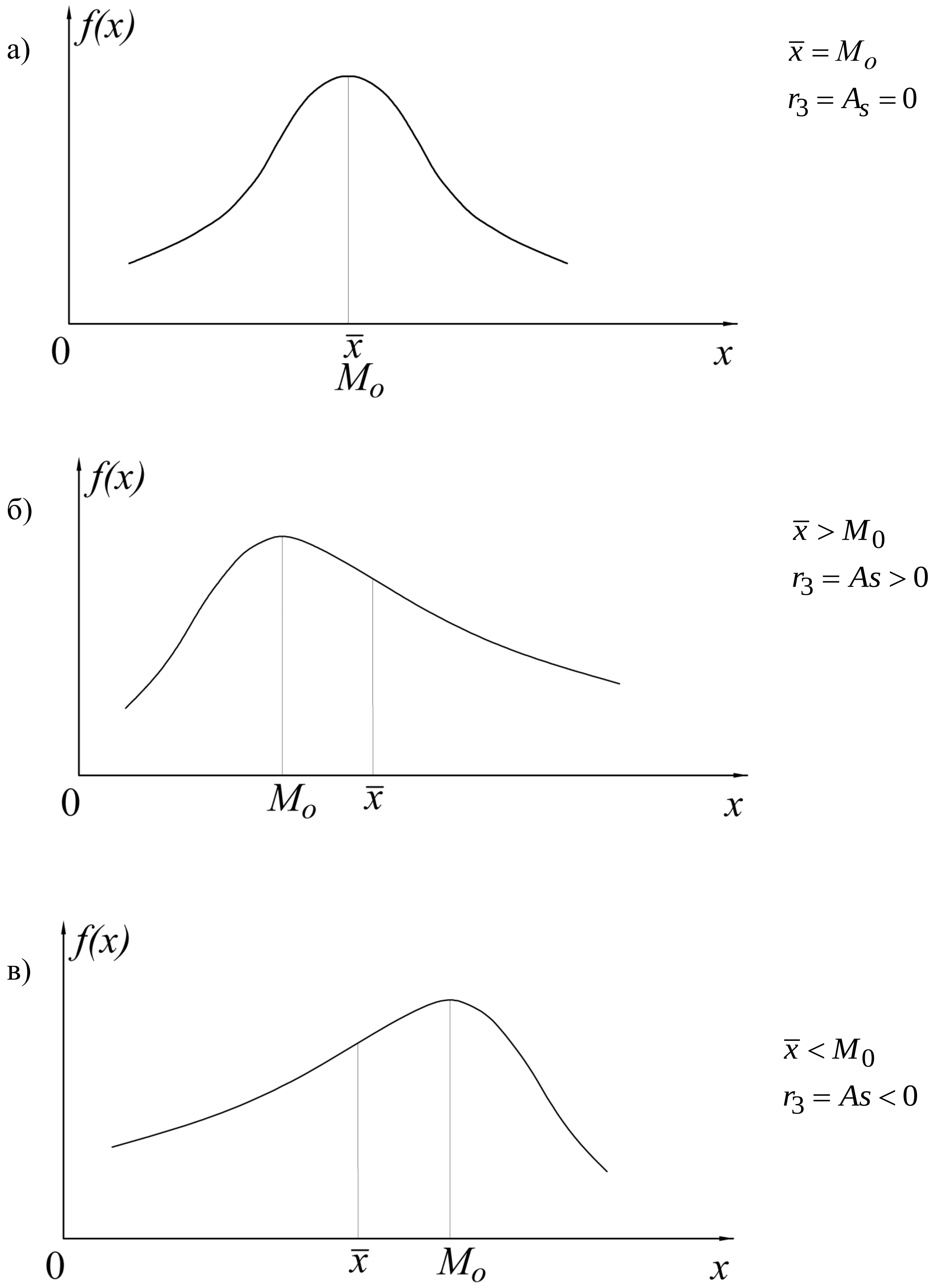
та *четвертого* (за умови l =4):

$$\hat{r}_4 = \frac{\hat{\mu}_4}{S_x^4} \tag{2.63}$$

порядків основних моментів розподілу.

Як відомо, оцінка третього основного моменту характеризує асиметрію кривої розподілу інтервальних частостей (або частот) і називається *коефіцієнтом асиметрії*: $\hat{r}_3 = As$. Крива розподілу має *правосторонню асиметрію* за умови $As > 0$, і *лівосторонню* – за умови $As < 0$. Вона є *симетричною* відносно центру розподілу, якщо $As = 0$ (рис. 2.2).

Як додатно, так і від'ємно асиметричні криві можуть бути слабко, помірно або різко асиметричними.



а) симетрична крива; б) крива з правосторонньою асиметрією;
в) крива з лівосторонньою асиметрією

Рис. 2.2 – Приклади кривих розподілу ймовірностей з різними значеннями коефіцієнта асиметрії

Заведено вважати асиметрію слабкою за умови $|As| \le 0,25$; помірною – за умови $0,25 < |As| \le 0,50$ та високою – за умови $|As| > 0,50$.

Крім асиметрії, крива розподілу, порівняно з кривою нормального розподілу, може бути *витягнутою* або *стислою*. Мірою цього є *коефіцієнт ексцесу E:*

$$E = \hat{r}_4 - 3. \qquad (2.64)$$

У першому випадку $E > 0$ і $r_4 > 3$; у другому – $E < 0$ і $r_4 < 3$.

За умов нормального розподілу $E = 0$ і $r_4 = 3$ (рис 2.3).

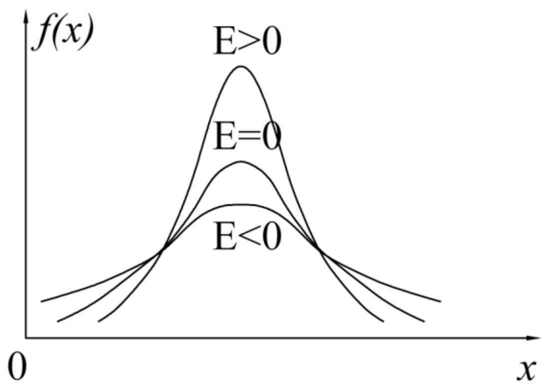


Рис. 2.3 – Приклади кривих розподілу ймовірностей з різними значеннями коефіцієнта ексцесу

Запитання для самоперевірки

1. Що називається «статистичною оцінкою параметра» генеральної сукупності?
2. Яким вимогам повинні відповідати статистичні оцінки параметрів, розрахованих30
3. на основі вибірок?
4. Яка статистична оцінка параметра генеральної сукупності називається «незсуненою»?
5. Яка статистична оцінка параметра генеральної сукупності називається «ефективною»?
6. Яка статистична оцінка параметра генеральної сукупності називається «умотивованою»?
7. Як розрахувати середнє значення випадкової величини на основі простої сукупності? згрупованого ряду?
8. Статистичною оцінкою якого моменту розподілу є середнє значення випадкової величини (вибірки)?

9. Який сенс дисперсії випадкової величини та з яким моментом розподілу вона має зв'язок?
10. З якою метою дослідник оцінює центральні моменти розподілу випадкової величини? Де ці статистичні оцінки використовуються?
11. Як розрахувати незсунену, ефективну та умотивовану оцінку дисперсії на основі простої вибірки? згрупованого ряду?
12. Підкресліть сенс коефіцієнта асиметрії для статистичних досліджень гідрометеорологічних процесів. Як розрахувати цей коефіцієнт на основі вибірки?
13. Як називають криву розподілу за умови AS = 0? AS > 0? AS < 0?
14. У чому сенс коефіцієнта ексцесу? Як розрахувати цей коефіцієнт на основі вибірки?
15. Як називається крива розподілу випадкової величини за умови E = 0? E > 0? E < 0?
16. Дати визначення «модального значення» випадкової величини. За якою формулою його можна розрахувати?
17. Дати визначення «медіани» випадкової величини. За якою формулою її можна розрахувати?
18. Який характер має крива розподілу випадкової величини за умови $\overline{x} < M_0$? Яких значень при цьому набуває коефіцієнт асиметрії?
19. Який характер має крива розподілу випадкової величини за умови $\overline{x} > M_0$? Яких значень при цьому набуває коефіцієнт асиметрії?
20. Який характер має крива розподілу випадкової величини за умови $\overline{x} = M_0$? Яких значень при цьому набуває коефіцієнт асиметрії?

3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

3.1 Загальні теоретичні положення щодо перевірки статистичних гіпотез

Перевірка статистичних гіпотез тісно поєднується з теорією оцінювання параметрів. У природознавстві, різних галузях техніки та економіки часто для з'ясування того чи іншого випадкового факту звертаються до висловлювання гіпотез, які можна перевірити статистично, тобто опираючись на результати спостережень у випадкових вибірках. Під *статистичними гіпотезами* розуміють такі гіпотези, котрі відносяться або до виду, або до окремих параметрів розподілу випадкової величини. Наприклад, статистичною є гіпотеза про те, що середня добова температура повітря має нормальний закон розподілу. Статистичною буде також гіпотеза про те, що ряди місячної кількісті опадів на сусідніх метеорологічних станціях суттєво не відрізняються.

Сформулюємо задачу статистичної перевірки гіпотези у загальному вигляді. Нехай $f\left(x,\theta\right)$ – закон розподілу випадкової величини X, який залежить від одного параметра θ. Припустимо, що необхідно перевірити гіпотезу про те, що $\theta=\theta_0$. Будемо називати цю *гіпотезу нульовою* і позначати її через H_0. Гіпотезу про те, що $\theta=\theta_1$ будемо називати *конкуруючою* і позначати її через H_1. Гіпотезу H_0 іноді називають *основною*, а гіпотезу H_1 – *альтернативною.* Отже, треба перевірити гіпотезу H_0 відносно конкуруючої гіпотези H_1 на основі вибірки, що складається з n незалежних спостережень над випадковою величиною X : $x_1, x_2, x_3, ..., x_n$.

Статистична сукупність розглядається як випадкова вибірка з генеральної сукупності. Отже, всю можливу множину (N) вибірок об'ємом n можна поділити на дві неперетинних підмножини (позначимо їх через U_1 і U_2), таких, що гіпотеза H_0 повинна бути відкинутою, якщо вибірка, яка розглядається, потрапляє до підмножини U_1, і прийнятою, якщо вибірка належить до підмножини U_2.

Зручніше використовувати не самі вибірки, а деякі статистичні параметри, що розраховуються на основі вибірок за визначеними правилами. Позначимо їх умовно q. Оскільки ці параметри є числами, а останні зображаються точками на числовій осі, підмножини U_1 та U_2 вибірок зводяться до двох підмножин точок числової осі або до двох одномірних областей W_1 і W_2. Область W_1 параметрів q називають *критичною областю*, а область W_2 – *областю припустимих значень*. Оскільки область W_2 складається з точок, які не увійшли до області W_1, то область W_1 однозначно визначає область W_2, і навпаки.

Виникає *питання* про те, *якими принципами треба керуватися при побудові критичної області* W_1? Ці принципи полягають у тому, що приймаючи чи відкидаючи гіпотезу H_0 можна припустити помилки двох видів.

Помилка першого роду полягає у тому, що нульова гіпотеза H_0 відкидається, тобто приймається гіпотеза H_1 тоді, коли в дійсності все ж таки вірною є гіпотеза H_0.

Помилка другого роду допускається тоді, коли приймається гіпотеза H_0 , у той час, коли вірною є гіпотеза H_1.

Імовірності помилок першого та другого роду однозначно визначаються вибором критичної області W_1. Умовимося для будь-якої критичної області W_1 позначати через α *імовірність помилки першого роду.* Її називають *рівнем значущості*. Критичну область W_1 відокремлює від області прийняття гіпотези H_0 критична точка ($q_{кр}$). Можуть розглядатися *правостороння, лівостороння, двостороння і симетрична двостороння критичні області.* Вони позначені на рис. 3.1.

Критичне значення параметра ($q_{кр}$) знаходиться за такої умови. Припустимо йдеться про *правосторонню критичну область* (рис. 3.1а). Нехай гіпотеза H_0 є вірною. Тоді ймовірність того, що гіпотеза H_0 відкидається, тобто, що робиться помилка I роду, дорівнює:

$$P\left(q > q_{кр}\right) = \alpha \,. \tag{3.1}$$

Для *лівосторонньої критичної області* (рис. 3.1б):

$$P\left(q < q_{кр}\right) = \alpha \,. \tag{3.2}$$

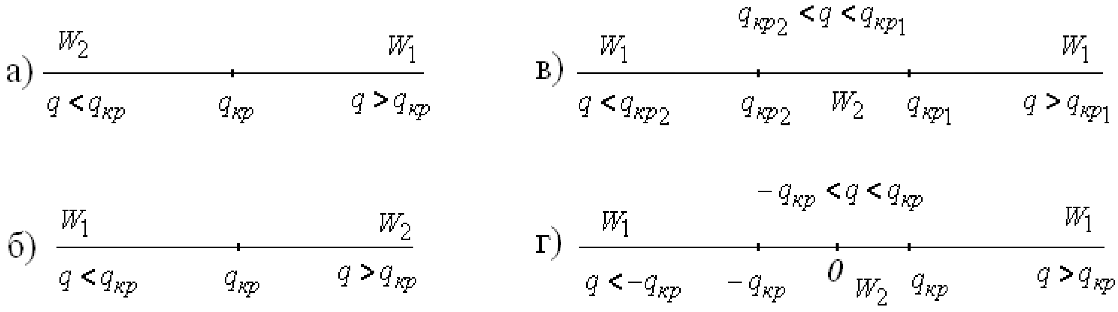
Для *двосторонньої критичної області* (рис. 3.1в):

$$P\left(q < q_{кр_2}\right) + P\left(q > q_{кр_1}\right) = \alpha \tag{3.3}$$

і для *двосторонньої симетричної області* (рис. 3.1г):

$$P\left(q < -q_{кр}\right) + P\left(q > q_{кр}\right) = \alpha \tag{3.4}$$

або $$P\left(q > q_{кр}\right) = \frac{\alpha}{2}. \tag{3.5}$$



а) правостороння; б) лівостороння; в) двостороння несиметрична; г) двостороння симетрична.

Рис. 3.1 – Види критичних областей

Отже можна сказати, що *при великій кількості вибірок частка помилкових рішень дорівнює* α *, якщо гіпотеза* H_0 *є вірною.*

Як вже підкреслювалося, параметр q формується за визначеними правилами в залежності від характеру задачі, яка розв'язується, та властивостей випадкових величин.

Для визначення *фактичного значення параметра* q використовуються такі теореми:

Теорема 1. Нехай незалежні величини $z_i: z_1, z_2, \ldots, z_n$ підпорядковуються нормальному закону розподілу. Тоді сума квадратів цих величин

$$\sum_{i=1}^{n} z_i^2 = \chi^2 \tag{3.6}$$

підпорядковується закону розподілу, який визначається щільністю ймовірності

$$f\left(\chi^2\right) = \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)}\left(\chi^2\right)^{\frac{\nu}{2}-1} e^{-\chi^2/2}, \qquad \text{за умови } \chi^2 \geq 0 \tag{3.7}$$

з *числом степенів вільності* ν. *Число степенів вільності є параметром цього розподілу. Число степенів вільності* – це кількість значень, функціонально не зв'язаних один з одним. Закон розподілу (3.7) називають χ^2 *– розподілом.* Він є однопараметричним.

Теорема 2. Нехай ми маємо дві незалежні випадкові величини U і V, які підпорядковуються χ^2 – розподілу з числами степенів вільності ν_1 і ν_2 відповідно. Тоді випадкова величина вигляду

$$\frac{U/\nu_1}{V/\nu_2} = F \tag{3.8}$$

підпорядковується розподілу, який описується щільністю ймовірності (3.9) і називається *законом Фішера – Снедекора:*

$$f(F) = \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}\frac{F^{\frac{\nu_1-2}{2}}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{\frac{\nu_1+\nu_2}{2}}}, \quad \textit{за умови } F \ge 0. \tag{3.9}$$

Цей розподіл є двопараметричним з параметрами ν_1 і ν_2.

Теорема 3. Нехай маємо дві незалежні випадкові величини U і V такі, що U – підпорядковується нормальному закону, а $V - \chi^2$ – розподілу з числом степенів вільності ν.

Тоді випадкова величина t

$$t = \frac{U}{\sqrt{\frac{V}{\nu}}} \tag{3.10}$$

підпорядковується розподілу з щільністю ймовірності (3.11):

$$f(t) = \frac{1}{\sqrt{\pi\nu}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty. \tag{3.11}$$

Рівняння (3.11) називається *розподілом Стьюдента*, який *має один параметр* ν.

Розглянемо ряд конкретних задач по перевірках статистичних гіпотез, які найчастіше реалізуються в гідрометеорологічних дослідженнях.

3.2 Задача перевірки гіпотези про однорідність членів статистичної сукупності

Відомо, що ряди спостережень, на основі яких розраховуються статистичні оцінки моментів розподілу, повинні мати деякі властивості (див. 1-й розділ).

Однією з вимог, яким повинна відповідати статистична сукупність, є *вимога однорідності її членів*. *Це означає, що всі члени вибірки з заданою ймовірністю повинні належати до однієї генеральної сукупності.* Однак, завдяки великій мінливості гідрометеорологічних процесів, фізичні параметри, що відбивають їх стан, можуть інколи змінюватися у дуже великих межах на фоні порівняно невеликих природних коливань. Ці різкі зміни фізичних параметрів обумовлюються аномальними гідрометеорологічними процесами. Значення фізичних параметрів, що відрізняються від їх середнього рівня (норми) отримали назву «*викидів*». «*Викиди*» спричиняють похибки при статистичному оцінюванні моментів розподілу випадкових величин, особливо центральних моментів. Щоб цього уникнути, треба «*викиди*» вилучати з вибірки, але тільки після перевірки статистичної гіпотези. Гіпотеза H_0 перевіряється на заданому рівні значущості α і формулюється таким чином: максимальне та мінімальне значення вибірки ($x_{екстр}$) належать до тієї ж генеральної сукупності, що і всі інші її члени.

Зрозуміло, якщо екстремальні значення статистичного ряду є однорідними, то й інші члени цієї вибірки також будуть підпорядковуватися одному закону розподілу.

Для перевірки цієї гіпотези використовують *критерій Стьюдента* t у вигляді:

$$t = \frac{\left| x_{екстр} - \bar{x} \right|}{S_x}, \tag{3.12}$$

який формується на основі відомих теорем (див. підрозділ 3.1).

Визначене за формулою (3.12) значення фактичного критерію Стьюдента (t) порівнюється з його критичним значенням – $t_{кр}(\alpha, \nu)$. Останнє отримують по таблицях (Додаток А) для визначеного дослідником рівня значущості (наприклад, α=0.05) і числа степенів вільності $\nu = n - 1$, де n – об'єм вибірки.

Якщо

$$t < t_{кр}(\alpha, \nu), \tag{3.13}$$

то гіпотеза H_0 не відкидається. Це означає, що екстремальний член сукупності, відносно якого існує підозра, у дійсності «викидом» не є, тобто його треба у вибірці залишити. Навпаки, якщо

$$t > t_{кр}(\alpha, \nu), \tag{3.14}$$

то гіпотеза H_0 відхиляється й приймається альтернативна гіпотеза H_1, про те, що $x_{екстр}$ на рівні значущості α не належить до тієї ж генеральної сукупності, що й інші члени вибірки, тобто є «викидом». У такому разі його слід вилучити з ряду значень перед розрахунками статистичних оцінок моментів розподілу випадкової величини.

Перевірку гіпотези H_0 треба повторювати до тих пір, поки на заданому рівні значущості члени статистичного ряду, що складають вибірку, будуть однорідними.

3.3 Перевірка статистичної гіпотези про однорідність двох нормально розподілених рядів

Два ряди випадкових величин називають статистично однорідними, якщо вони на рівні значущості α належать до однієї генеральної сукупності (підпорядковуються одному закону розподілу).

Будемо вважати, що ряди

$$X: \quad x_1, \quad x_2, \ \ldots, \ x_i, \quad x_n \tag{3.15}$$

і

$$Y: \quad y_1, \quad y_2, \ \ldots, \ y_i, \quad y_m \tag{3.16}$$

з об'ємами відповідно n та m підпорядковуються нормальному закону.

У такому випадку щільність імовірності нормального розподілу для випадкової величини X визначається рівнянням:

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{(x - m_x)^2}{2\sigma_x^2}\right], \tag{3.17}$$

а рівняння (3.18) визначає щільність імовірності нормального розподілу для випадкової величини Y:

$$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{(y - m_y)^2}{2\sigma_y^2}\right]. \tag{3.18}$$

Для таких випадкових величин *закони розподілу будуть однаковими, якщо їх відповідні параметри (масштабу та форми) будуть дорівнювати один одному: $m_x = m_y$; $\sigma_x^2 = \sigma_y^2$.*

Але дослідник аналізує окремі вибірки величин X та Y, а на їх основі можливо отримати лише статистичні оцінки параметрів розподілу: $\hat{m}_x = \overline{x}$, $\hat{m}_y = \overline{y}$, $\hat{\sigma}_x^2 \Rightarrow S_x^2$, $\hat{\sigma}_y^2 \Rightarrow S_y^2$.

У такому *разі ряди випадкових величин X та Y будуть однорідними*, тобто будуть на рівні значущості α належати до однієї генеральної сукупності, *якщо різниці між статистичними оцінками параметрів їх розподілів,* тобто S_x^2 і S_y^2 з одного боку та $\overline{x}$ і $\overline{y}$ *– з іншого,носитимуть випадковий характер* або, інакше кажучи, *не будуть значущими*.

Як випливає з розглянутих вище умов, основна гіпотеза H_0 є складною і дослідження однорідності випадкових рядів полягає у перевірці двох статистичних гіпотез на заданому рівні значущості: гіпотези H_0' про незначущість різниць статистичних оцінок дисперсій та гіпотези H_0'' про незначущість розбіжностей між середніми значеннями випадкових величин X та Y.

а) Перевірка гіпотези H_0' про незначущість різниць між статистичними оцінками дисперсій.

Для перевірки гіпотези H_0' використовується критерій Фішера-Снедекора. Фактичне значення цього критерію сформовано на основі відомих теорем (див. підрозділ 3.1) і визначається формулою:

$$F = \frac{S_x^2}{S_y^2} \quad \text{за умови} \quad S_x^2 > S_y^2 \tag{3.19}$$

(у чисельник рівності (3.19) треба поставити більшу з цих статистичних оцінок дисперсій).

Оскільки закон розподілу випадкової величини F є двопараметричним з параметрами − числами степенів вільності $\nu_1 = n-1$ і $\nu_2 = m-1$, α – задає дослідник, то критичне значення критерію Фішера-Снедекора є функцією трьох величин: ν_1, ν_2 та α, тобто $F_{кр}(\alpha, \nu_1, \nu_2)$. Його отримують по таблицях за визначеними характеристиками (Додаток Б).

Якщо статистична оцінка дисперсії випадкової величини Y (S_y^2) більша за S_x^2, фактичне значення критерію F формують таким чином:

$$F = \frac{S_y^2}{S_x^2}. \tag{3.20}$$

Відповідно з цим $\nu_1 = m-1$ і $\nu_2 = n-1$. *Завжди число степенів вільності* ν_1 *розраховується по об'єму тієї вибірки, статистична оцінка дисперсії котрої більша.*

Критичне значення критерію Фішера – $F_{кр}(\alpha,\nu_1,\nu_2)$ порівнюють зі значенням, отриманим за формулою (3.19) або (3.20).

Якщо

$$F < F_{кр}(\alpha,\nu_1,\nu_2), \tag{3.21}$$

то гіпотеза H_0' про незначущість розбіжностей між статистичними оцінками дисперсій S_x^2 і S_y^2 не відхиляється.

Якщо виявиться навпаки:

$$F > F_{кр}(\alpha,\nu_1,\nu_2), \tag{3.22}$$

то гіпотеза H_0' відхиляється й приймається альтернативна гіпотеза H_1' про те, що статистичні оцінки дисперсій рядів X та Y розрізняються статистично значуще. Відхилення гіпотези H_0' призводить і до відхилення основної гіпотези H_0.

Якщо гіпотеза H_0' не відхиляється, це ще не означає, що ряди X та Y – однорідні. Треба перевірити іншу гіпотезу – H_0''.

б) Перевірка статистичної гіпотези H_0'' *про незначущість розбіжностей між середніми значеннями двох вибірок.*

Гіпотеза H_0'' формулюється так: на рівні значущості α середні значення випадкових величин X і Y розрізняються статистично незначуще. Перевірка цієї гіпотези проводиться за допомогою критерію Стьюдента. Його фактичне значення формується на основі ряду теорем (див. підрозділ 3.1) і визначається рівнянням:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{S_x^2(n-1)+S_y^2(m-1)}{n+m-2}\left(\dfrac{1}{n}+\dfrac{1}{m}\right)}}. \tag{3.23}$$

Модуль у чисельнику формули (3.23) застосовується для того, щоб розглядати лише додатні значення параметра t. Крім того, під знаком радикалу вводиться нормуючий множник $\left(\dfrac{1}{n}+\dfrac{1}{m}\right)$.

Розподіл випадкової величини t є *однопараметричним*. Це число степенів вільності ν, яке знаходять за формулою (3.24):

$$\nu = n + m - 2. \quad (3.24)$$

Нагадаємо, що n та m – об'єми вибірок випадкових величин X та Y. Критичні значення параметра Стьюдента – $t_{кр}(\alpha, \nu)$ для рівня значущості α і числа степенів вільності ν знаходять по таблицях (Додаток А).

Якщо виконується нерівність

$$t < t_{кр}(\alpha, \nu), \quad (3.25)$$

то гіпотеза H_0'' не відхиляється.

У протилежному випадку, за умови

$$t > t_{кр}(\alpha, \nu), \quad (3.26)$$

приймається гіпотеза H_1'' про статистичну значущість розбіжностей між середніми значеннями рядів X та Y ($\bar{x}$ і $\bar{y}$).

Гіпотеза H_0 про однорідність двох рядів випадкових величин не відхиляється лише у тому випадку, коли не відхиляються гіпотези і H_0', і H_0''. Якщо будь-яка з них відхиляється, то відхиляється й основна гіпотеза H_0.

3.4 Перевірка гіпотези про однорідність двох рядів випадкових величин за допомогою непараметричного критерію Вілкоксона

Критерії Фішера-Снедекора (F) і *Стьюдента* (t) називають *параметричними*. Вони використовуються тоді, коли відомо, що випадкові величини X та Y є нормально розподіленими. У гідрометеорологічних дослідженнях ця умова не завжди виконується. Дуже часто ряди випадкових величин підпорядковуються законам розподілу, що відрізняються від нормального закону. В такому випадку для перевірки гіпотези H_0 про однорідність рядів таких випадкових величин застосовують *непараметричні критерії*. Ці критерії *треба використовувати і тоді, коли інформація про закон розподілу випадкових величин X та Y відсутня.*

Одним із *непараметричних критеріїв* є *критерій Вілкоксона*. Цей критерій використовується в *двох варіантах*: перший базується на підрахунках *числа інверсій*, інший – *ранговий*.

3.4.1 Інверсійний критерій Вілкоксона

Якщо ми маємо прості статистичні ряди випадкових величин у вигляді (3.15) і (3.16), то дослідження однорідності таких вибірок треба розпочинати з *операції ранжування*. Тобто випадкові величини, що належать до цих вибірок, розташовують у загальну послідовність у порядку збільшення (або зменшення) їх значень.

Якщо будь-якому значенню x *у загальній послідовності передує деяке значення* y, *то кажуть, що ця пара утворює інверсію*. Позначимо сумарну кількість інверсій через U_x. Відповідно, якщо будь-якому значенню y в загальній послідовності передує деяке значення x, то ця пара також утворює інверсію (U_y).

Відомо, що в однорідних рядах, кожний з котрих має не менше 10 членів, число інверсій розподіляється приблизно за нормальним законом з параметрами:

$$m_U = \frac{m \cdot n}{2} \tag{3.27}$$

і

$$\sigma_U^2 = \frac{m \cdot n}{12}\left(m + n + 1\right), \tag{3.28}$$

де

n і m – кількість членів у першій і другій вибірках.

Нульову гіпотезу H_0 сформулюємо так: ряди випадкових величин X та Y належать до однієї генеральної сукупності на рівні значущості α.

Для перевірки цієї гіпотези необхідно знайти межі допустимих значень U, що відділяють область прийняття гіпотези H_0 від критичної області, яка, як зрозуміло, є двосторонньою. Якщо значення критерію, що отримане за даними спостережень, попаде до області прийняття гіпотези, то гіпотеза H_0 на цьому рівні значущості не відхиляється. В іншому випадку береться альтернативна гіпотеза H_1.

Область прийняття гіпотези H_0 визначається виразом:

$$m_U - t_{кр}\left(\alpha, \nu\right)\sigma_U \le U \le m_U + t_{кр}\left(\alpha, \nu\right)\sigma_U, \tag{3.29}$$

а *критична* – нерівностями:

$$U_{кр_1} < m_U - t_{кр}\left(\alpha, \nu\right)\sigma_U, \tag{3.30}$$

$$U_{кр_2} > m_U + t_{кр}(\alpha, \nu)\sigma_U. \tag{3.31}$$

У нерівностях (3.29)-(3.31) $\sigma_U = \sqrt{\sigma_U^2}$ – середній квадратичний відхил числа інверсій; $t_{кр}(\alpha, \nu)$ – критерій Стьюдента для рівня значущості α і числа степенів вільності $\nu = m + n - 2$, який знаходять по таблицях.

3.4.2 Ранговий варіант критерію Вілкоксона

Будемо вважати, що хоча б одна з вибірок (X чи Y) складає більше 25 членів.

За допомогою рангового критерію Вілкоксона перевірка статистичної гіпотези H_0 про однорідність двох рядів випадкових величин починається з того, що:

- розташовують члени обох вибірок в один об'єднаний ряд (в порядку збільшення їх значень);
- далі знаходять в об'єднаному ряду суму порядкових номерів (рангів) членів вибірки, об'єм якої менший. Позначимо її через $W_{виб.}$;
- потім визначають нижню критичну точку за формулою:

$$W_{кр.н}\left(m, n, \alpha/2\right) = \frac{(m+n+1)\cdot m - 1}{2} - Z_{кр}\sqrt{\frac{m \cdot n \cdot (m+n+1)}{12}}, \tag{3.32}$$

де n і m – об'єми вибірок випадкових величин X та Y ;

α – рівень значущості;

$Z_{кр}$ – визначається по таблицях функції Лапласа за правилом:

$$\Phi\left(Z_{кр}\right) = \frac{1-\alpha}{2}. \tag{3.33}$$

Верхня критична точка розраховується за формулою:

$$W_{кр.в} = (m+n+1)\cdot m - W_{кр.н}. \tag{3.34}$$

Гіпотеза H_0 на заданому рівні значущості α не відкидається, якщо

$$W_{кр.н} < W_{виб.} < W_{кр.в}. \tag{3.35}$$

За умов $W_{виб.} < W_{кр.н}$ або $W_{виб.} > W_{кр.в}$ береться альтернативна гіпотеза H_1 про те, що ряди X та Y - неоднорідні.

Запитання для самоперевірки.

1. Що розуміють під терміном «статистична гіпотеза»?
2. Які критичні області використовуються при статистичних дослідженнях?
3. Якими принципами треба керуватися при побудові тієї чи іншої критичної області?
4. Який сенс рівня значущості? Дати визначення «рівня значущості» для правосторонньої критичної області?
5. Як називається «ймовірність помилки I-го роду»?
6. Що розуміють під терміном «довірча ймовірність»?
7. За якими правилами та теоремами формується фактичний критерій (q) до перевірки тієї чи іншої статистичної гіпотези?
8. На якому принципі втілюється перевірка статистичної гіпотези?
9. Яким вимогам повинні відповідати вибірки, по яких розраховуються статистичні оцінки параметрів генеральної сукупності?
10. Які члени статистичного ряду називаються «однорідними»? Які значення вибірки перевіряються на однорідність?
11. Які значення вибірки отримали назву «викиди»? Якщо вибірка має «викиди», як необхідно діяти, щоб розрахувати статистичні оцінки моментів розподілу випадкової величини?
12. Як «викиди» впливають на розрахунки статистичних оцінок параметрів генеральної сукупності?
13. Від чого залежить формування фактичного критерію Стьюдента для перевірки членів статистичного ряду на однорідність?
14. Які статистичні оцінки моментів розподілу випадкової величини необхідно розрахувати, щоб отримати фактичний критерій Стьюдента для знаходження «викидів» у статистичній сукупності?
15. Від яких величин залежить критичне значення критерію Стьюдента для перевірки членів статистичного ряду на однорідність?
16. Які статистичні сукупності називаються «однорідними» та які критерії використовуються для перевірки статистичної гіпотези про однорідність двох рядів випадкових величин?
17. Чому статистична гіпотеза перевірки двох рядів на однорідність за допомогою параметричних критеріїв є складною?
18. Яким чином формується фактичний критерій для перевірки статистичної гіпотези про незначущість різниць оцінок дисперсій, отриманих по двох сукупностях однієї і тієї ж гідрометеорологічної величини?
19. Яким чином формується фактичний критерій для перевірки статистичної гіпотези про незначущість розбіжностей між середніми значеннями, що розраховані по двох рядах однієї і тієї ж гідрометеорологічної величини?

20. Як визначаються критичні точки критерію Фішера для перевірки статистичної гіпотези про незначущість розбіжностей між оцінками дисперсій двох рядів?
21. Як визначаються критичні точки критерію Стьюдента для перевірки статистичної гіпотези про незначущість розбіжностей між середніми значеннями двох рядів?
22. В якому випадку дві статистичні сукупності випадкових величин, що підпорядковуються нормальному закону, будуть однорідними?
23. За яких умов використовуються параметричні критерії для перевірки двох рядів випадкових величин на однорідність?
24. За яких умов використовуються непараметричні критерії для перевірки двох рядів випадкових величин на однорідність?
25. В яких випадках використовується інверсійний критерій Вілкоксона?
26. Поясніть терміни «інверсія», «сума інверсій» в задачах перевірки двох рядів випадкових величин на однорідність?
27. Який ряд значень випадкової величини називають «ранжованим»?
28. Якими нерівностями визначаються критичні області при застосуванні критерію Вілкоксона?
29. За яких умов при використанні інверсійного критерію Вілкоксона два ряди випадкових величин будуть однорідними?
30. За яких умов застосовується ранговий критерій Вілкоксона для перевірки на однорідність двох рядів випадкових величин?
31. Що таке «сума рангів»?
32. Як розраховується ранг, якщо однакові значення належать різним вибіркам?
33. За яких умов при використанні рангового критерію Вілкоксона два ряди випадкових величин будуть однорідними?
34. Як в прикладних цілях використовується результат перевірки гіпотези H_0 на заданому рівні значущості про однорідність двох рядів випадкових величин?
35. Як фізично інтерпретувати результат перевірки двох рядів на однорідність у випадку їх неоднорідності?
36. Як фізично інтерпретувати результат перевірки двох рядів на однорідність у випадку їх однорідності?

4. ЗАКОНИ РОЗПОДІЛУ ГІДРОМЕТЕОРОЛОГІЧНИХ ВЕЛИЧИН ТА АЛГОРИТМИ ЇХ ДОСЛІДЖЕННЯ

4.1 Поняття про закон розподілу та форми його представлення

Законом розподілу випадкової величини називають будь-яку відповідність між можливими значеннями випадкової величини та їх імовірностями. Для випадкової величини закон розподілу є вичерпною характеристикою. Знання закону, якому підпорядковується та чи інша гідрометеорологічна величина, дає можливість методично правильно організувати дослідження статистичної структури цих величин, які є основними ознаками кліматичного чи гідрологічного режиму регіону, що досліджується. Підібравши закон розподілу до статистичного ряду (вибірки), можна розрахувати ймовірність того, що випадкова величина знаходиться у заданому інтервалі або ймовірність того, що випадкова величина прийме значення менше (більше) деякого конкретного числа з області значень цієї випадкової величини.

Дослідження законів розподілу гідрометеорологічних величин має велике практичне значення. Як уже зазначалося неодноразово, в основних ланках кліматичної системи (атмосфері та гідросфері) постійно відбуваються зміни їх фізичного стану, а кількісні характеристики цього стану, такі, наприклад, як температура повітря чи води, атмосферний тиск, хмарність, вологість повітря, кількість опадів, річний стік та інші розглядаються як випадкові величини.

Задача дослідника полягає у тому, щоб серед множини випадкових подій чи явищ виявити закономірності, відкинувши несуттєві події. А це можна зробити шляхом побудови моделей фізичних параметрів, які висвітлюють властивості цих випадкових величин, які, як відомо, містяться у законі розподілу.

Таким чином, вичерпною характеристикою будь-якого випадкового процесу є закон розподілу, знання якого дає можливість правильно протулмачити сенс того чи іншого статистичного моменту та на основі цього методично правильно організувати вивчення гідрометеорологічних особливостей регіону, що досліджується.

У більшості випадків закони розподілу гідрометеорологічних величин неможливо визначити апріорно тільки шляхом аналізу відомих фізичних властивостей. Тип розподілу та його параметри визначаються шляхом статистичної обробки експериментальних даних. Найбільш поширеним є метод їх групування, при якому вся множина значень випадкової величини розділяється на ряд неперетинних часткових інтервалів, а потім підраховується кількість значень ряду, що потрапили до кожного часткового інтервалу. Але таким шляхом, як було показано в підрозділі 1.2, будується і згрупований ряд. Отже *згрупований ряд має сенс емпіричного розподілу випадкової величини,* графічним зображенням якого є гістограма або полігон.

Таким чином, вивчення особливостей статистичної структури гідрометеорологічних величин базується на інформації, в якості якої виступають статистичні ряди (вибірки), що сформовані за результатами вимірювань та спостережень.

Згрупований ряд, як емпіричний розподіл, апроксимують аналітичним виразом, який відбиває властивості генеральної сукупності. У зв'язку з цим, основним етапом статистичного аналізу гідрометеорологічної інформації є підбір закону розподілу за даними статистичної сукупності. Ця задача розв'язується шляхом апроксимації емпіричного розподілу таким теоретичним законом, який би у визначеному сенсі найкращим чином відповідав емпіричному розподілу. Але, як би добре, на підставі відомих властивостей закону розподілу, не була б підібрана теоретична крива чи то нормального розподілу, чи то розподілів Пірсона, Пуассона або інших відомих, про властивості яких піде мова у цьому розділі пізніше, між нею та емпіричним (статистичним) розподілом неминучі деякі розбіжності. Тому обов'язково після розрахунків теоретичних частот (за відповідними формулами, в залежності від закону розподілу) проводять перевірку гіпотези про міру розбіжності між емпіричними та теоретичними частотами. Розбіжності між цими частотами можуть носити як випадковий характер, так і бути статистично значущими. Останнє вказує на те, що підібрана теоретична крива не відповідає даному емпіричному розподілу. Щоб з'ясувати ці питання використовують так звані «*критерії згоди»*. І тільки після перевірки відповідної статистичної гіпотези на заданому рівні значущості можна зробити висновок про успішність апроксимації статистичного розподілу теоретичним законом.

Закон розподілу випадкової величини X може визначатися у вигляді *функції розподілу* $F(x)$ або *щільності ймовірності* $f(x)$. Дуже часто його представляють сукупністю *інтервальних ймовірностей* p_i або *інтервальних теоретичних частот* $\tilde{m}_i$.

4.1.1 Функція розподілу та її властивості

Функція розподілу є найбільш загальною формою аналітичного представлення закону розподілу. Вона зазначає ймовірність того, що випадкова величина X набирає значення, меншого від фіксованого дійсного числа x. Для неперервної випадкової величини це визначення має вигляд:

$$F(x) = P(X < x) \ . \tag{4.1}$$

Імовірність того, що $X < x$ залежить від x, отже $F(x)$ є функцією від x. Тому $F(x)$ і називається *функцією розподілу.*

Неперервна випадкова величина має неперервну чи кусково-неперервну функцію розподілу; графік цієї функції має форму плавної кривої (рис. 4.1).

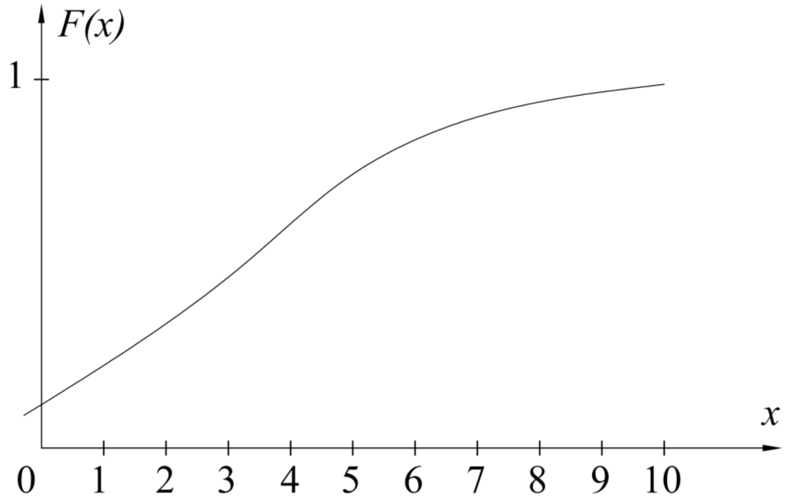


Рис. 4.1 – Функція розподілу неперервної випадкової величини

Для дискретної випадкової величини X, яка може набувати значеннь $x_1, x_2, ..., x_i, ...x_n$, функція розподілу має вигляд:

$$F(x) = \sum_{x_i < x} P(X = x_i), \tag{4.2}$$

де нерівність $x_i < x$ під знаком суми означає, що підсумування поширюється на всі значення x_i, менші за x.

Функція розподілу має стрибок у тих точках, де випадкова величина приймає конкретні значення. В інтервалах між значеннями випадкової величини функція $F(x)$ є сталою. Сума всіх стрибків функції розподілу дорівнює одиниці. Графік функції розподілу дискретної випадкової величини є розривна східчаста ламана (рис. 4.2).

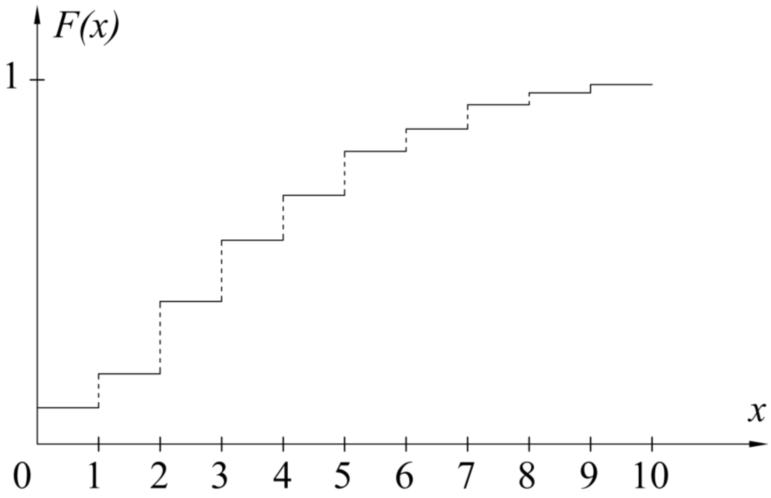


Рис. 4.2 – Функція розподілу дискретної випадкової величини

Розглянемо *загальні властивості функції розподілу*.

1. Функція розподілу $F\left(x\right)$ є невід'ємною функцією з областю значень

$$0 \le F\left(x\right) \le 1. \tag{4.3}$$

Ця властивість випливає з визначення функції розподілу як імовірності здійснювання події $X < x$.

2. На мінус нескінченності функція розподілу $F\left(x\right)$ дорівнює нулю, а на плюс нескінченності – функція розподілу дорівнює одиниці, тобто

$$F\left(-\infty\right) = 0; \quad F\left(+\infty\right) = 1. \tag{4.4}$$

Ця властивість стає очевидною при геометричній інтерпретації функції розподілу (див. рис. 4.2). Якщо точка x необмежено пересувається ліворуч, то попадання X ліворуч від x у границі стає неможливою подією. Тому можна вважати, що ймовірність цієї події прагне до нуля, тобто

$$F\left(-\infty\right) = \lim_{x \to -\infty} F\left(x\right) = 0. \tag{4.5}$$

При необмеженому пересуванні точки x праворуч, попадання випадкової точки X праворуч від x у границі стає вірогідною подією. Тому можна вважати, що ймовірність цієї події прагне до одиниці, тобто

$$F\left(+\infty\right) = \lim_{x \to +\infty} F\left(x\right) = 1. \tag{4.6}$$

3. Імовірність попадання випадкової величини в інтервал $[\alpha, \beta]$ дорівнює різниці значень функції розподілу на кінцях цього інтервалу

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha). \tag{4.7}$$

4. Функція розподілу випадкової величини є неспадною функцією

$$F(\beta) \geq F(a), \text{ якщо } \beta > \alpha. \tag{4.8}$$

Очевидно,

$$P(X > x) = 1 - F(x). \tag{4.9}$$

Функція розподілу, як і будь-яка ймовірність, є безрозмірною величиною.

Неперервну випадкову величину можна задати не тільки функцією розподілу, але й щільністю ймовірностей. Розглянемо цю форму подання закону розподілу випадкової величини.

4.1.2 Щільність імовірності та її властивості

Нехай випадкова величина X визначається функцією розподілу $F(x)$. Згідно з третьою властивістю функції розподілу ймовірність попадання цієї величини в елементарний інтервал $[x; x + \Delta x]$ дорівнює:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x). \tag{4.10}$$

Розділимо обидві частини на Δx:

$$\frac{P(x < X < x + \Delta x)}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x}. \tag{4.11}$$

Вважаючи, що функція розподілу неперервна, перейдемо до границі за умови $\Delta x \to 0$. Тоді отримаємо похідну від функції розподілу, яка й називається *щільністю ймовірності*

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}. \tag{4.12}$$

Крива, що зображує щільність імовірності випадкової величини $f(x)$, називається *кривою розподілу*. Чисельні приклади кривих розподілу будуть розглядатися у наступних підрозділах.

Тепер можна привести визначення *неперервної випадкової величини*: випадкова величина X називається неперервною, якщо її функція розподілу $F(x)$ неперервна на всій осі OX, а щільність розподілу $f(x)$ існує всюди, за винятком скінченної кількості точок.

Розглянемо *властивості щільності ймовірності.*

1. Щільність імовірності є функцією невід'ємною :

$$f(x) \geq 0. \tag{4.13}$$

Ця властивість випливає безпосередньо з визначення цієї функції як похідної від неспадної функції $F(x)$.

2. Імовірність попадання неперервної випадкової величини X до інтервалу $[\alpha;\beta]$ дорівнює :

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx. \tag{4.14}$$

3. Функцію розподілу можна виразити через щільність імовірності за формулою:

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{4.15}$$

4. Інтеграл у нескінченних межах від щільності ймовірності дорівнює одиниці:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{4.16}$$

Геометрично це означає, що площа між кривою розподілу та віссю абсцис дорівнює одиниці.

4.1.3 Представлення закону розподілу в термінах інтервальних теоретичних частот

Якщо область значень гідрометеорологічної величини розділяється на k часткових інтервалів довжиною c, то, знаючи щільність імовірності, яка притаманна цій випадковій величині, можна знайти, використовуючи відомі властивості цієї функції, ймовірність того, що випадкова величина належить до будь-якого i-го часткового інтервалу з його серединою $\tilde{x}_i$. Ця ймовірність визначається формулою:

$$p_i = P\left[\tilde{x}_i - \frac{c}{2} < x < \tilde{x}_i + \frac{c}{2}\right] = \int_{\tilde{x}_i - \frac{c}{2}}^{\tilde{x}_i + \frac{c}{2}} f(x)\,dx \qquad \left(i = \overline{1,k}\right). \tag{4.17}$$

Розрахувавши ці ймовірності для кожного часткового інтервалу, отримаємо закон розподілу гідрометеорологічної величини, що досліджується, у вигляді сукупності значень цієї величини на серединах градацій $\tilde{x}_i$ та інтервальних імовірностей p_i:

$$\begin{cases} \tilde{x}_1\,, \tilde{x}_2\,, ..., \; \tilde{x}_i\,, ..., \; \tilde{x}_k \\ p_1\,, p_2\,, ..., p_i\,, ..., p_k \end{cases} \quad \left(i = \overline{1,k}\right). \tag{4.18}$$

Як випливає з виразу (4.18), даний закон розподілу має структуру, подібну до структури згрупованого ряду (1.11). Маючи це на увазі, а також пам'ятаючи, що *частість* $\hat{p}_i$ *є незсуненою, ефективною та умотивованою оцінкою ймовірності* p_i, приходимо до висновку, що *згрупований ряд* (1.11) *можна розглядати як емпіричний розподіл цієї величини. Графічним аналогом емпіричного розподілу є гістограма або полігон,* що відповідає цьому згрупованому ряду.

За допомогою простої рівності

$$\tilde{m}_i = p_i \cdot n, \tag{4.19}$$

де n – об'єм вихідного статистичного ряду, до якого добирається теоретичний закон розподілу, сукупність інтервальних імовірностей перетворюється у відповідну сукупність теоретичних інтервальних частот $\tilde{m}_i$:

$$\begin{cases} \tilde{x}_1, \; \tilde{x}_2, ..., \; \tilde{x}_i, ..., \; \tilde{x}_k \\ \tilde{m}_1, \tilde{m}_2, ..., \; \tilde{m}_i, ..., \; \tilde{m}_k \end{cases} \quad \left(i = \overline{1,k}\right). \tag{4.20}$$

Таким чином, можна отримати закон розподілу випадкової величини X, який визначається сукупністю інтервальних теоретичних частот та значень випадкової величини на серединах градацій.

4.2 Алгоритм дослідження закону розподілу

Оскільки практика ставить перед природничими науками задачу пізнання законів, які б дозволили керувати процесами, що відбуваються на Землі в її географічних сферах. Фізичні та хімічні процеси, що виникають і розвиваються в складових кліматичної системи (атмосфера, гідросфера, літосфера, кріосфера, біосфера) вивчаються багатьма науками. Але гідрометеорологічні науки, які входять до складу наук про Землю, відділяються певними особливостями напряму розвитку та методологією досліджень.

Зупинимося на основних положеннях щодо апроксимації емпіричних розподілів відомими теоретичними законами.

Процес дослідження закону розподілу складається з таких етапів:

- по-перше, від простої статистичної сукупності випадкової величини, що розглядається, необхідно перейти до згрупованого ряду (див. розділ 1) і на основі зовнішнього вигляду отриманого емпіричного розподілу, який зображується полігоном чи гістограмою, з урахуванням статистичних оцінок моментів розподілу та деяких (в залежності від закону) допоміжних статистик, з заданою ймовірністю сформулювати гіпотезу H_0 про закон розподілу, яким можна апроксимувати даний емпіричний розподіл;
- по-друге, на основі статистичних оцінок моментів розподілу гідрометеорологічної величини, що досліджується, за відповідними формулами знаходять оцінки параметрів вибраного теоретичного розподілу;
- на третьому етапі для кожної градації за основними рівняннями визначеного теоретичного закону розраховують теоретичні інтервальні частоти $\tilde{m}_i$. У підрозділах 4.4 та 4.5 розглянемо більш детально властивості та методику розрахунків інтервальних теоретичних частот для тих законів, які найчастіше використовуються при статистичних дослідженнях гідрометеорологічних процесів.

  Для з'ясування якісних розбіжностей між емпіричними m_i та теоретичними $\tilde{m}_i$ частотами в кожному інтервалі доцільно побудувати полігони даного емпіричного та отриманого теоретичного розподілів;
- на завершальному етапі за визначеною методикою, що буде розглядатися в підрозділі 4.3, роблять кількісну оцінку розбіжності між емпіричними та теоретичними інтервальними частотами, тобто треба на заданому рівні значущості α (див. розділ 3) за допомогою відомих критеріїв з'ясувати статистично значущі чи незначущі ці розбіжності. Якщо розбіжності статистично незначущі, то вони пояснюються тільки випадковими обставинами, пов'язаними, наприклад, з обмеженою кількістю спостережень. Якщо розбіжності суттєві (статистично значущі), то пов'язані вони з тим, що теоретичний закон розподілу випадкової величини з заданою ймовірністю не відповідає емпіричному розподілу. Задачу апроксимації емпіричного розподілу теоретичним законом (за умови прийняття основної гіпотези H_0) слід завершувати аналітичним представленням отриманого теоретичного розподілу, використовуючи для цього або щільність ймовірності $f(x)$, або функцію розподілу $F(x)$, або інтервальні теоретичні частоти $\tilde{m}_i$, оскільки вони, як відомо, пов'язані з інтервальною ймовірністю співвідношенням (4.19).

4.3 Перевірка статистичної гіпотези про відповідність

емпіричного та теоретичного розподілів

Реалізація алгоритму апроксимації емпіричного розподілу відомим теоретичним законом приводить до отримання інтервальних теоретичних частот за визначеними формулами в залежності від теоретичного закону.

Отже, для кожної i -ої середини градації можна записати:

$$\begin{array}{l} \tilde{x}_1,\ \tilde{x}_2,\ \tilde{x}_3, \ldots,\ \tilde{x}_i, \ldots,\ \tilde{x}_k \\ m_1, m_2, m_3, \ldots, m_i, \ldots,\ m_k \qquad\qquad i = \overline{1,k}, \\ \tilde{m}_1, \tilde{m}_2, \tilde{m}_3, \ldots, \tilde{m}_i, \ldots, \tilde{m}_k \end{array} \tag{4.21}$$

де m_i та $\tilde{m}_i$ – емпірична та інтервальна теоретична частоти відповідно.

Виникає питання, яка міра розбіжності між емпіричними та теоретичними інтервальними частотами нас задовольняє?

Дійсно, як би добре не була б підібрана теоретична крива, між нею та статистичним розподілом завжди будуть деякі розбіжності. Задача дослідника полягає в тому, щоб з'ясувати статистично значущі чи незначущі ці розбіжності. *Якщо розбіжності статистично незначущі,* то вони пояснюються тільки випадковими обставинами, пов'язаними, наприклад, з обмеженою кількістю спостережень. *Якщо розбіжності суттєві (статистично значущі),* то пов'язані вони з тим, що теоретичний закон розподілу випадкової величини не відповідає емпіричному розподілу.

Відповіді на ці запитання отримаємо після перевірки статистичної гіпотези. Для цього, перш за все, треба визначити критерій, за допомогою якого втілюється перевірка гіпотези H_0 відносно альтернативної гіпотези H_1. Гіпотеза H_0 формулюється таким чином: розбіжності між емпіричними m_i і теоретичними $\tilde{m}_i$ частотами є несуттєвими на рівні значущості α.

Для її перевірки використовують так звані *«критерії згоди».* Одним з них є *критерій Пірсона* χ^2, який вважається досить зручним. Він визначається формулою:

$$\chi^2 = \sum_{i=1}^{k'} \frac{\left(m_i - \tilde{m}_i\right)^2}{\tilde{m}_i}\ . \tag{4.22}$$

Застосування цього критерію для перевірки узгодженості між емпіричними та теоретичними частотами є досить простим, тому що, як випливає з рівності (4.22), він складається з різниць між цими частотами, що й визначає міру розбіжності між ними.

Далі отримане за формулою (4.22) значення критерію Пірсона χ^2 порівнюють з критичним – $\chi^2_{кр}(\alpha,\nu)$ (Додаток В). Останнє залежить від рівня значущості α та числа степенів вільності ν. Рівень значущості α визначає дослідник. Число степенів вільності залежить від типу розподілу і визначається рівнянням:

$$\nu = k' - l, \tag{4.23}$$

де

k' – кількість часткових інтервалів, для яких $m_i \geq 5$;

l – кількість лінійних зв'язків відносно інтервальних частот m_i, що використовуються при статистичному оцінюванні моментів розподілу, на основі яких розраховуються оцінки параметрів того чи іншого теоретичного закону.

Наведемо *числа степенів вільності для визначення* $\chi^2_{кр}(\alpha,\nu)$ при перевірках гіпотез для тих законів розподілу, якими найчастіше апроксимуються гідрометеорологічні процеси:

- *Нормальний розподіл:* $\nu = k' - 3$.
- *I тип розподілів Пірсона:* $\nu = k' - 5$.
- *II тип розподілів Пірсона:* $\nu = k' - 4$.
- *III тип розподілів Пірсона:* $\nu = k' - 4$.
- *Гамма-розподіл:* $\nu = k' - 3$.
- *Розподіл Пуассона:* $\nu = k' - 2$.

Рішення відносно прийняття гіпотези H_0 або альтернативної гіпотези H_1 залежить від того, в яку область попаде χ^2.

Для *правосторонньої критичної області* (3.1)

$$P\left[\chi^2 > \chi^2_{кр}(\alpha,\nu)\right] = \alpha. \tag{4.24}$$

Тому нерівність

$$\chi^2 < \chi^2_{кр}(\alpha,\nu) \tag{4.25}$$

визначає область прийняття гіпотези H_0, а критична область визначається нерівністю:

$$\chi^2 > \chi^2_{кр}(\alpha,\nu). \tag{4.26}$$

У першому випадку, якщо отримаємо нерівність (4.25), гіпотеза H_0 про незначущість розбіжностей між емпіричними (m_i) та теоретичними ($\tilde{m}_i$) частотами і, таким чином, про відповідність емпіричного розподілу теоретичному, не відкидається. За умови нерівності (4.26) – приймається гіпотеза H_1 й це означає, що теоретичний розподіл, який вибрано для апроксимації емпіричного розподілу, з заданою ймовірністю не відповідає останньому.

4.4 Закони розподілу Пірсона

4.4.1 Загальна теорія розподілів Пірсона

Нехай ми маємо статистичний ряд деякої випадкової величини X об'ємом n. Побудуємо згрупований ряд ($\tilde{x}_i, m_i$; $i = \overline{1,k}$), де

$\tilde{x}_i$ – середина часткового інтервалу;

m_i – емпірична інтервальна частота,

і перейдемо до безрозмірної випадкової величини

$$z = \frac{x - \hat{x}}{c}, \tag{4.27}$$

де

$\hat{x}$ – деяка статистика, що буде визначена пізніше;

c – довжина часткового інтервалу.

Розглянемо диференціальне рівняння

$$\frac{dy}{dz} = y\frac{z+b}{c_2 z^2 + c_1 z + c_0}, \tag{4.28}$$

в якому

b, c_0, c_1, c_2 – дійсні числа.

Якщо змінна z у ньому є випадковою величиною, пов'язаною з вихідною випадковою величиною X рівністю (4.27), і якщо коефіцієнти c_0, c_1, c_2 є статистиками випадкової величини X, такими, що

$$c_0 = -\sigma^2 \frac{S+1}{S-2}; \tag{4.29}$$

$$c_1 = -b = -\frac{r_3 \sigma}{2} \frac{(S+2)}{(S-2)}; \tag{4.30}$$

$$c_2 = \frac{1}{S-2};\qquad (4.31)$$

$$\sigma = \frac{\sigma_x}{c};\qquad (4.32)$$

$$S = \frac{6(r_4 - r_3^2 - 1)}{3r_3^2 - 2r_4 + 6}\ ;\qquad (4.33)$$

σ_x – середній квадратичний відхил величини X, r_3 та r_4 – її третій та четвертий основні моменти, то розв'язок диференціального рівняння (4.28) являє собою сукупність законів розподілу Пірсона. Тобто, *розподіл Пірсона* – це сім'я розподілів імовірностей, щільність яких $y = P(z)$ задовольняє диференціальному рівнянню (4.28). Відповідні графіки $y = P(z)$, що відображають залежність щільності ймовірності від z, називають *кривими Пірсона* взагалі.

Розподіли Пірсона класифікують у залежності від значення коефіцієнтів b, c_0, c_1, c_2 та області змінення z. Сім'я розподілів Пірсона складається з 12 типів. Будь-який розподіл Пірсона однозначно визначається своїми першими чотирма початковими моментами

$$v_l = \int_{-\infty}^{\infty} z f(z) dz \quad (l = \overline{1,4}).\qquad (4.34)$$

Розглянемо теорію розподілів Пірсона більш детально.
Розв'язок рівняння (4.28), очевидно, має такий вигляд :

$$y = y_0 e^{\int \frac{z+b}{c_2 z^2 + c_1 z + c_0} dz},\qquad (4.35)$$

де y_0 - довільна стала.

Як відомо, інтеграл у правій частині рівності (4.35) залежить від характеру коефіцієнтів квадратного трьохчлена

$$Q(z) = c_2 z^2 + c_1 z + c_0\ .\qquad (4.36)$$

Підставимо до нього значення коефіцієнтів (4.29)-(4.31) і перетворимо його у рівняння. Розв'язок квадратного рівняння дає такі значення коренів трьохчлена (4.36):

$$z_{1,2} = \sigma \frac{r_3(S+2) \pm t}{4},\qquad (4.37)$$

де

$$t = 4\sqrt{(S+1)(1-æ)}\,, \qquad (4.38)$$

а

$$æ = -\frac{r_3{}^2 (S+2)^2}{16(S+1)}\ . \qquad (4.39)$$

Після математичих перетворень рівність (4.38) приймає вигляд:

$$t = \frac{16|S+1|}{|r_3||S+2|}\sqrt{æ(æ-1)}\ . \qquad (4.40)$$

Із рівності (4.40) випливає, що значення коренів квадратного тричлена (4.36) залежать від величини

$$u = æ(æ-1)\,. \qquad (4.41)$$

Рівняння (4.41), якщо його доповнити до повного квадрата

$$u + \frac{1}{4} = (æ - \frac{1}{2})^2, \qquad (4.42)$$

є рівнянням параболи з вершиною в точці з координатами $(\frac{1}{2};-\frac{1}{4})$. Графік функції (4.41) наводиться на рис. 4.3.

Як виходить з формули (4.41) і графіка функції $u = f(æ)$, область значень æ може бути розподіленою на три підобласті, в кожній з яких функція u має один і той же знак. У відповідності до цього, статистика (4.40) відноситься до множини дійсних або уявних чисел. У залежності від цього, набуває певного вигляду інтеграл в експоненті розв'язку (4.35) і, таким чином, сам розв'язок. Оскільки кожен із розв'язків диференціального рівняння (4.28) співвідноситься з визначеним типом розподілу Пірсона, кожному з підобластей значень æ теж відповідає той чи інший *тип розподілів Пірсона*, а саме:

– області $æ < 0$ відповідає I тип;
– області $0 < æ < 1$ відповідає IV тип;
– області $æ > 1$ відповідає VI тип;
– значенню æ =0 відповідають:

$$\begin{cases} \text{II тип за умови } r_4 < 3\ ; \\ \text{нормальний розподіл за умови } r_4 = 0; \\ \text{VII тип за умови } r_4 > 3\ ; \end{cases}$$

– значенню $æ = 1$ відповідає V тип;
– області $-\infty < æ < \infty$ відповідає III тип.

Отже, III тип розподілів Пірсона може спостерігатися за будь-яких значень статистики $æ$.

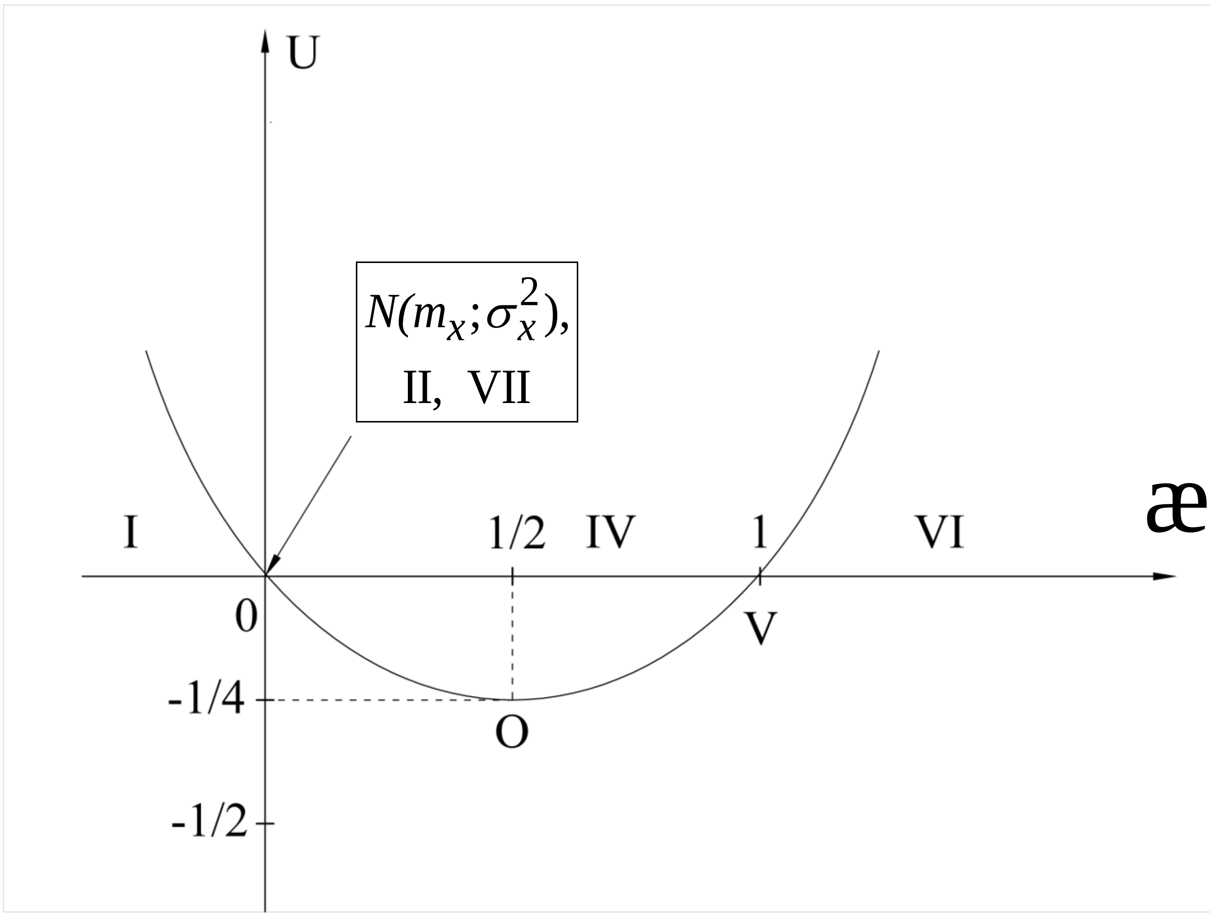


Рис. 4.3 – $æ$-діаграма розподілів Пірсона [9]

Якщо для кожної з областей значень $æ$ отримати щільність імовірності $y = f(z)$ для того чи іншого типу розподілів Пірсона і проінтегрувати її в межах i-го часткового інтервалу, то будемо мати *інтервальну ймовірність*

$$p_i = P[z_1 + (i-1)c < z < z_1 + ic] = \int_{z_1+(i-1)c}^{z_1+ic} y(z)dz \quad , \tag{4.43}$$

де

$$z_1 = \frac{x_1 - \hat{x}}{c} \quad - \tag{4.44}$$

значення лівої межі області значень змінної x у новій системі координат.

З іншого боку:

$$p_i = \frac{\tilde{m}_i}{n}, \quad (4.45)$$

де

$\tilde{m}_i$ – *теоретична частота* для i-го часткового інтервалу;

n – загальний об'єм сукупності випадкової величини.

Таким чином, розподіли Пірсона можуть розглядатися і в термінах інтервальних теоретичних частот $\tilde{m}_i$.

Статистика $\hat{x}$ визначається формулою:

$$\hat{x} = \bar{x} - \frac{r_3 \sigma_x}{2} \frac{(S+2)}{(S-2)} \ . \quad (4.46)$$

4.4.2 Нормальний розподіл як частинний випадок розподілів Пірсона

Як було зазначено, при æ=0, коли r_4=3 (у точці æ=0, очевидно, $r_3 = 0$), розв'язок диференціального рівняння (4.28) відповідає нормальному розподілу. Обґрунтуємо цей факт. Для цього знайдемо значення статистик S, c_2, c_1, c_0 при зазначених вище основних моментах r_3 та r_4

$$S = \frac{6(3-1)}{-2 \cdot 3 + 6} = \infty; \quad (4.47)$$

$$c_2 = \lim_{S \to \infty} \frac{1}{S-2} = 0; \quad (4.48)$$

$$c_1 = -b = \lim_{S \to \infty} \left[-\frac{r_3 \sigma}{2} \frac{(S+2)}{(S-2)} \right] = -\frac{r_3 \sigma}{2} = 0; \quad (4.49)$$

$$c_0 = -\lim_{S \to \infty} \left[\sigma^2 \frac{S+1}{S-2} \right] = -\sigma^2. \quad (4.50)$$

Отже, інтеграл у показнику степеня розв'язку (4.35) набуває значення:

$$\int \frac{z+b}{c_2 z^2 + c_1 z + c_0} dz = -\int \frac{z}{\sigma^2} dz = -\frac{z^2}{2\sigma^2}. \quad (4.51)$$

Оскільки, як видно з рівності (4.46), і враховуючи формули (4.27) та (4.32) і те, що $\bar{x}$ є оцінкою m_x, маємо:

$$\int \frac{z+b}{c_2 z^2 + c_1 z + c_0} dz = -\frac{(x-m_x)^2}{2\sigma_x{}^2}. \tag{4.52}$$

Будемо вважати, що довільна стала y_0 дорівнює

$$y_0 = \frac{1}{\sqrt{2\pi}\sigma_x}. \tag{4.53}$$

Тоді щільність імовірності та функція розподілу нормального закону для випадкової величини X визначаються рівняннями:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left\{-\frac{(x-m_x)^2}{2\sigma_x^2}\right\}, \tag{4.54}$$

$$-\infty < x < \infty$$

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(x)\, dx. \tag{4.55}$$

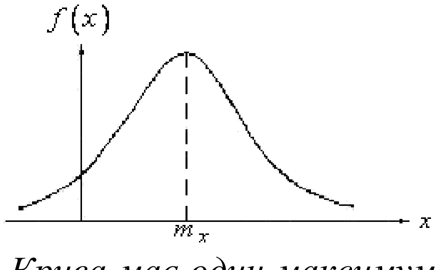
Параметрами цього закону є математичне сподівання m_x (*параметр масштабу*) і дисперсія σ_x^2 (*параметр форми*).

4.4.2.1 Властивості нормального розподілу

Нормальному закону підпорядковується багато гідрометеорологічних величин. Наприклад, температура повітря, парціальний тиск водяної пари, атмосферний тиск біля земної поверхні та у вільній атмосфері і деякі інші метеорологічні та гідрологічні ряди спостережень. Тому нормальний розподіл (закон Гаусса) часто використовується в статистичних дослідженнях атмосфери та гідросфери.

Визначимо *основні властивості* нормального розподілу.

1. *Графік кривої розподілу (щільності ймовірності)* для випадкової величини X має вигляд:



Вона симетрична відносно ординати, яка проходить через точку, що дорівнює *математичному сподіванню* m_x, а воно є *центром нормального розподілу.*

2. *Крива має один максимум* за умови $x = m_x$ (*розподіл одномодальний*). В точці максимуму щільність ймовірності дорівнює:

$$f(x)_{\max} = \frac{1}{\sigma_x \sqrt{2\pi}}. \tag{4.56}$$

3. За умови $|x| \to \infty$ вітки кривої асимптотично наближаються до осі OX.
4. Згідно з симетричністю кривої розподілу, *математичне сподівання, мода й медіана нормально розподіленої випадкової величини співпадають:* $m_x = M_o = Me$.
5. *Крива розподілу має дві точки перегину* з координатами:

$$\left[(m_x - \sigma_x); \frac{1}{\sigma_x \sqrt{2\pi e}}\right]; \left[(m_x + \sigma_x); \frac{1}{\sigma_x \sqrt{2\pi e}}\right]. \tag{4.57}$$

6. *Непарні центральні моменти нормального розподілу дорівнюють нулю*. За допомогою *рекурентного співвідношення*

$$\mu_l = (l-1)\sigma_x^2 \mu_{l-2}, \quad l = 2,3,4,... \tag{4.58}$$

є можливість знайти центральні моменти вищих порядків через моменти нижчих порядків. Крім того маємо:

$\mu_2 = \sigma_x^2$; $\mu_4 = 3\sigma_x^4$; $\mu_6 = 15\sigma_x^6$ і т.п.

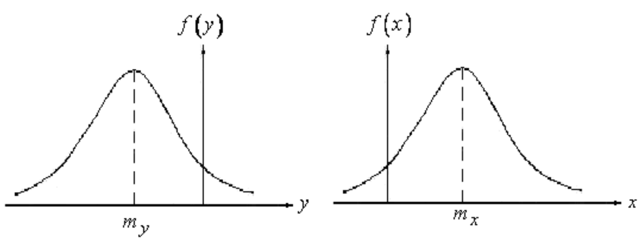
7. *Коефіцієнти асиметрії та ексцесу дорівнюють нулю*:

$$r_3 = As = \frac{\mu_3}{\sigma_x^3} = 0; \tag{4.59}$$

$$E = r_4 - 3 = \frac{\mu_4}{\sigma_x^4} - 3 = \frac{3\sigma_x^4}{\sigma_x^4} - 3 = 0. \tag{4.60}$$

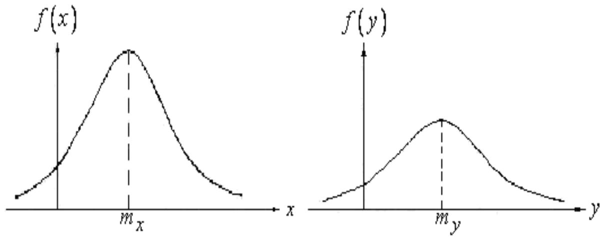
Стає зрозумілою важливість обчислювання цих коефіцієнтів для емпіричних рядів, оскільки один (As) є *мірою скосу*, а інший (E) – характеризує *стислість (чи витягнутість) кривої даного розподілу порівняно з кривою нормального розподілу.*

8. *Форма кривої нормального розподілу не змінюється зі зміною математичного сподівання* m_x *(параметра масштабу).*



При зменшенні (чи збільшенні) математичного сподівання графік щільності ймовірності зсувається ліворуч (праворуч) по осі OX.

9. *При змінюванні дисперсії* σ_x^2 *змінюється форма кривої розподілу випадкової величини X.* Зі збільшенням дисперсії максимальна ордината, що визначається рівністю (4.56), зменшується, а зі зменшенням σ_x^2 – збільшується. *Згідно з властивістю щільності ймовірності, площа, що обмежена кривою розподілу і віссю абсцис, дорівнює одиниці.* Тому при зростанні дисперсії крива розподілу розтягується вздовж осі OX.



10. *Якщо випадкова величина X нормально розподілена, для неї справедливими є три співвідношення:*

$$P\left[m_x-\sigma_x<x<m_x+\sigma_x\right]=0.6827; \tag{4.61}$$

$$P\left[m_x-2\sigma_x<x<m_x+2\sigma_x\right]=0.9545; \tag{4.62}$$

$$P\left[m_x-3\sigma_x<x<m_x+3\sigma_x\right]=0.9973. \tag{4.63}$$

Графічно співвідношення (4.61)-(4.63) представлені на рис. 4.4.

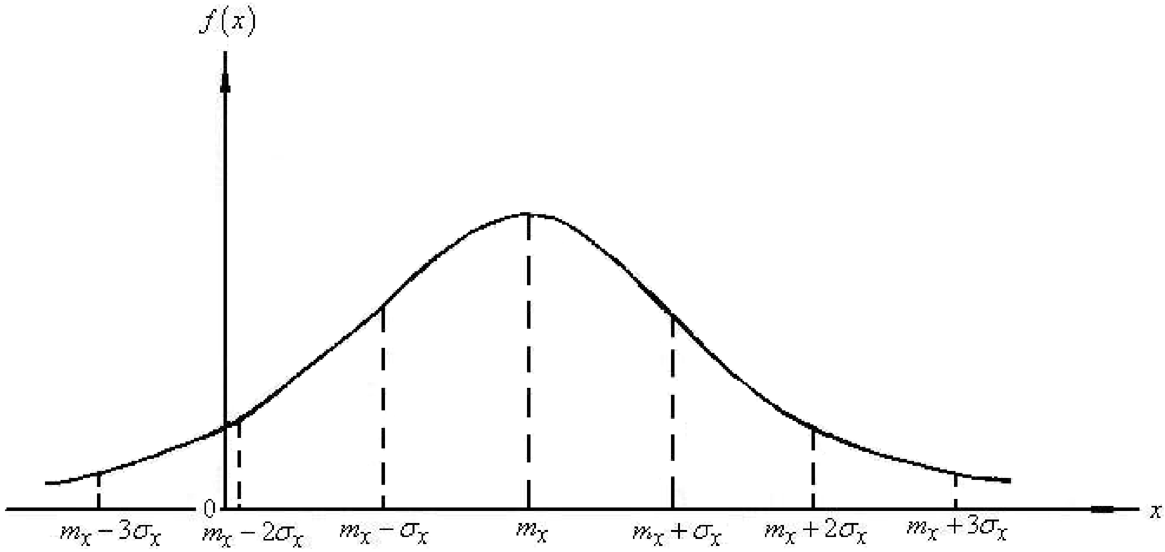


Рис. 4.4 - Щільність імовірності нормального розподілу [1]

З рівності (4.63) та рис. 4.4 виходить, що практично розсіяння нормально розподіленої випадкової величини *X* укладається на інтервалі $m_x \pm 3\sigma_x$. Імовірність того, що значення випадкової величини *X* виходить за цей інтервал, дуже мала і дорівнює 0.0027. Така подія може вважатися практично неможливою. На проведеному міркуванні основується *правило трьох сигм*, яке формулюється таким чином: *якщо випадкова величина має нормальний розподіл, то відхил цієї величини від математичного сподівання за абсолютною величиною не перевищує трьох середніх квадратичних відхилів.* Але, якщо остання властивість виконується, це *ще не означає*, що випадкова величина X підпорядковується нормальному розподілу.

Термін «*нормальний розподіл»* застосовується в умовному сенсі як загально прийнятий в літературі термін. *Головна особливість нормального закону полягає у тому, що він є граничним законом, до якого наближаються інші закони розподілу.*

4.4.2.2 Методика розрахунків інтервальних теоретичних частот нормального розподілу

Оскільки теоретичний закон розподілу можна представляти не тільки функцією розподілу або щільністю ймовірності, а й в термінах інтервальних теоретичних частот, тому ознайомимося з цією методикою більш детально.

Для нормального розподілу розрахунок інтервальної теоретичної частоти для i-го часткового інтервалу можна проводити за двома методами. Але спочатку треба перейти від меж i-го часткового інтервалу вихідної гідрометеорологічної величини (x_{i-1} і x_{i+1}) до відповідних нових центрованих і нормованих величин для лівої межі (t_{i-1}) і правої межі (t_{i+1}) кожної градації. Для цього використовуються такі формули:

$$t_{i-1} = \frac{x_{i-1} - \overline{x}}{S_x} \qquad \left(i = \overline{1,k}\right); \tag{4.64}$$

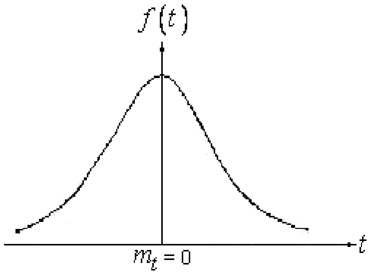
$$t_{i+1} = \frac{x_{i+1} - \overline{x}}{S_x} \qquad \left(i = \overline{1,k}\right), \tag{4.65}$$

де S_x – оцінка середнього квадратичного відхилу випадкової величини X;

$\overline{x}$ – середнє значення, яке є оцінкою математичного сподівання випадкової величини *X*.

Один із шляхів обчислення інтервальних теоретичних частот нормального розподілу передбачає використання значень *щільності ймовірності нормованого нормального розділу* $f(t)$:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. \tag{4.66}$$



Крива розподілу (4.66) *симетрична відносно математичного сподівання центрованої і нормованої випадкової величини, яке є центром цього розподілу* ($m_t = 0$). Дисперсія такої величини дорівнює одиниці $\left(\sigma_t^2 = 1\right)$.

Значення функції (4.66) можна отримати по таблицях (Додаток Г).

Для розрахунку інтервальної теоретичної частоти i-го часткового інтервалу використовується формула:

$$\tilde{m}_i = \frac{nc}{S_x} f\left(t_i\right) \qquad \left(i = \overline{1,k}\right), \tag{4.67}$$

де

t_i – безрозмірне значення випадкової величини на середині i-го часткового інтервалу, що розраховується за формулою:

$$t_i = \frac{t_{i+1} + t_{i-1}}{2}. \tag{4.68}$$

Функція $f(t)$ *є функцією парною*: $f\left(-t\right) = f\left(t\right)$.

Інший шлях обчислення інтервальних теоретичних частот нормального розподілу полягає у використанні *інтеграла ймовірності* $\Phi\left(t\right)$, який має вигляд:

$$\Phi\left(t\right) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt. \tag{4.69}$$

Значення цього інтеграла для різних t затабульовані (Додаток Д).

За допомогою інтеграла ймовірності інтервальна ймовірність для i-го часткового інтервалу розраховується так:

$$p_i = P\left(t_{i-1} < t < t_{i+1}\right) = \frac{1}{2}\left[\Phi\left(t_{i+1}\right) - \Phi\left(t_{i-1}\right)\right] \qquad \left(i = \overline{1,k}\right), \tag{4.70}$$

а теоретичні інтервальні частоти за формулою:

$$\tilde{m}_i = n\, p_i \quad \left(i = \overline{1,k}\right), \tag{4.71}$$

де, як і раніше, n – об’єм вибірки (статистичного ряду).

Треба мати на увазі, що *інтеграл імовірності є непарною функцією*, тобто $\Phi\left(-t\right) = -\Phi\left(t\right)$.

4.4.3 Перший тип розподілів Пірсона

У відповідності до загальної теорії розподілів Пірсона (пункт 4.4.1) *перший тип спостерігається за умови від'ємного значення статистики капа* $(æ < 0)$. Вона розраховується за такою формулою:

$$æ = -\frac{r_3^2 (S+2)^2}{16(S+1)}. \tag{4.72}$$

Статистику S отримаємо як:

$$S = \frac{6\left(r_4 - r_3^2 - 1\right)}{3r_3^2 - 2r_4 + 6}, \tag{4.73}$$

де

r_3 та r_4 – третій та четвертий основні моменти розподілу випадкової величини.

Для першого типу розподілів Пірсона основне рівняння має вигляд:

$$\tilde{m}_i = \tilde{m}_0 \left(1 + \frac{z_i}{l_1}\right)^{q_1} \left(1 - \frac{z_i}{l_2}\right)^{q_2}. \tag{4.74}$$

У рівнянні (4.74) $q_1,\ q_2,\ l_1,\ l_2,\ \tilde{m}_0$ – параметри цього розподілу, а змінна z_i – безрозмірна величина, яка дорівнює:

$$z_i = \frac{\tilde{x}_i - \hat{x}}{c}, \tag{4.75}$$

де

$$\hat{x} = \overline{x} - \frac{S_x \cdot r_3 (S+2)}{2(S-2)}. \tag{4.76}$$

Параметри форми q_1, q_2 визначаються як:

$$\left.\begin{matrix} q_1 \\ q_2 \end{matrix}\right\} = \frac{1}{2}\left[(S-2) \mp S(S+2)\frac{r_3}{t}\right], \tag{4.77}$$

де t – статистика, що розраховується за допомогою рівняння:

$$t = 4\sqrt{(S+1)(1-æ)}. \tag{4.78}$$

Справедлива рівність (4.79):

$$q_1 + q_2 = S - 2. \tag{4.79}$$

Параметри масштабу l_1, l_2 оцінюють за формулами (4.80) та (4.81):

$$l_1 = \frac{q_1 l}{S-2} \ ; \tag{4.80}$$

$$l_2 = \frac{q_2 l}{S-2} \ , \tag{4.81}$$

де l_1 – ліва межа, а l_2 – права межа визначення функції (4.74).
Область існування функції (4.74) можна визначити також рівняннями:

$$l = l_1 + l_2 \quad (4.82); \qquad l = \frac{\sigma t}{2} \quad (4.83), \quad \text{де } \sigma = \frac{S_x}{c}. \quad (4.84)$$

Формула (4.84) свідчить про те, що σ є безрозмірне середнє квадратичне відхилення, оскільки c – довжина часткового інтервалу.

Параметр $\tilde{m}_0$ – *теоретична частота за умови* $z_i = 0$ і розраховується за допомогою рівняння (4.85):

$$\tilde{m}_0 = \frac{n\, q_1^{q_1} q_2^{q_2}}{l(S-2)^{S-2}} \frac{\Gamma(q_1 + q_2 + 2)}{\Gamma(q_1 + 1)\Gamma(q_2 + 1)} \ , \tag{4.85}$$

де

n – загальний об'єм вибірки;

$\Gamma(x)$ – гамма-функція відповідного аргументу, яка знаходиться по таблицях (Додаток Е).

Вигляд кривих розподілу Пірсона першого типу (рис. 4.5) залежить від *параметрів форми* q_1 і q_2, а *параметри* l_1 та l_2 *визначають межі області визначення функції* (4.74). У відповідності з цим, *виділяють 9 кривих,* які необхідно враховувати при апроксимації емпіричних розподілів першим типом розподілів Пірсона.

Алгоритм апроксимації емпіричного розподілу І типом розподілів Пірсона складається з 4-х основних етапів.

Відомо, що при дослідженнях законів розподілу випадкових величин, *перш за все,* необхідно від простої статистичної сукупності $X: x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ перейти до згрупованого ряду та графічно представити отриманий емпіричний розподіл. Розрахувати статистичні оцінки першого початкового $(\hat{\nu}_1 = \bar{x})$, другого центрального $(\hat{\mu}_2)$ моментів розподілу, отримати незсунену, ефективну і умотивовану оцінку дисперсії $\left(S_x^2\right)$ та середнього квадратичного відхилення (S_x) випадкової величини, що досліджується. Оцінити коефіцієнти асиметрії $(As = \hat{r}_3)$ та ексцесу (E). Отримати статистики S, æ, t і, якщо для æ виконується невірність æ < 0, то формулюють гіпотезу на заданому рівні значущості (розділ 3) про можливість апроксимації емпіричного розділу I типом розподілів Пірсона.

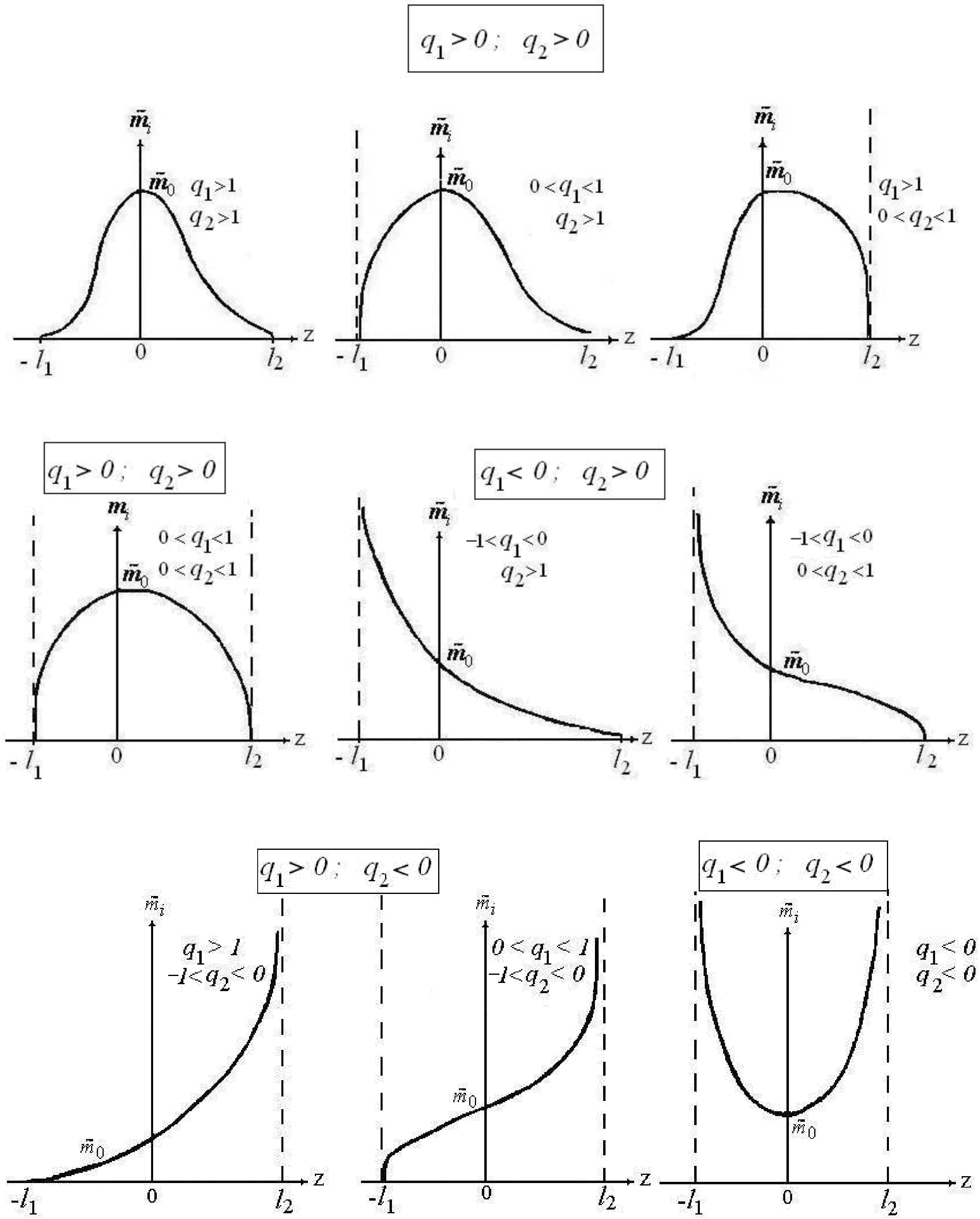


Рис. 4.5 – Криві розподілу першого типу Пірсона за умов різних значень параметрів форми $(q_1, \ q_2)$ та параметрів масштабу $(l_1, \ l_2)$ [1]

На *другому етапі* за відповідними формулами оцінюють параметри цього розподілу, а саме $q_1,\ q_2,\ l_1,\ l_2,\ \tilde{m}_0$ та необхідні для цього статистики.

На *третьому етапі* розраховують інтервальні теоретичні частоти за допомогою основного рівняння I типу (4.74). Але спочатку треба отримати статистику $\hat{x}$, використовуючи рівняння (4.76) і формулу (4.75) перейти від розмірних значень середин часткових інтервалів $\tilde{x}_i$ до безрозмірних величин z_i $\left(i=\overline{1,k}\right)$.

Розраховані для кожної градації інтервальні теоретичні частоти $\tilde{m}_i$ порівнюють з відповідними емпіричними частотами m_i і на основі отриманих характеристик (вираз (4.86)) будують полігони цих розподілів та з'ясовують спочатку якісні розбіжності між m_i та $\tilde{m}_i$:

$$\begin{array}{l} \tilde{x}_1,\ \ \tilde{x}_2,\ \ldots,\ \ \tilde{x}_i,\ \ldots,\ \tilde{x}_k \\ m_1,\ m_2,\ \ldots,\ m_i,\ \ldots,\ m_k\ . \\ \tilde{m}_1,\ \tilde{m}_2,\ \ldots,\ \tilde{m}_i,\ \ldots,\ \tilde{m}_k \end{array} \tag{4.86}$$

Метод порівняння інтервальних емпіричних та теоретичних частот за допомогою відомих критеріїв (див. підрозділ 4.3) дає змогу отримати остаточну відповідь на можливість (з заданою ймовірністю) апроксимації емпіричного розподілу першим типом розподілів Пірсона. Тому на *четвертому етапі* треба перевірити статистичну гіпотезу про відповідність емпіричного та теоретичного розподілів. Оскільки перевірка цієї гіпотези відносно типу розподілу є самостійною задачею, алгоритм її рішення розглядався окремо (див. підрозділ 4.3).

4.4.4 Другий тип розподілів Пірсона

У відповідності до загальної теорії розподілів Пірсона, *другий тип спостерігається за умови значення статистики капа, яка дорівнює нулю* (æ $=0$). Вона розраховується за формулою (4.72). При цьому $r_3=0$ (*розподіл є симетричним відносно центру розподілу)*, а $r_4<3$.

Основне рівняння другого типу розподілів Пірсона має вигляд:

$$\tilde{m}_i=\tilde{m}_0\left[1-\frac{{z_i}^2}{\tilde{l}^2}\right]^q, \tag{4.87}$$

де $q,\ \tilde{l},\ \tilde{m}_0$ – параметри II типу розподілів Пірсона.

Параметр форми q визначається рівнянням (4.88):

$$q = \frac{5r_4 - 9}{2(3 - r_4)}. \tag{4.88}$$

Параметр масштабу $\tilde{l}$ розраховується за формулою (4.89):

$$\tilde{l} = \sigma\sqrt{\frac{2r_4}{3 - r_4}}. \tag{4.89}$$

Параметр $\tilde{m}_0$ *(теоретична частота за умови* $z_i = 0$) можна отримати за допомогою рівняння (4.90):

$$\tilde{m}_0 = \frac{n\Gamma(2q + 2)}{\tilde{l}\; 2^{2q+1}\{\Gamma(q+1)\}^2}, \tag{4.90}$$

де $\Gamma(x)$ – гамма функція відповідного аргументу. Її значення для заданого x знаходяться по таблицях (Додаток Е).

Якщо порівняти рівняння (4.74) та (4.87), то видно, що другий тип за певних умов, а саме $q_1 = q_2 = q$ та $l_1 = l_2 = \tilde{l}$ отримаємо з першого типу розподілів Пірсона.

Другий тип розподілів Пірсона об'єднує три типи кривих (рис. 4.6). Оскільки цьому типу відповідає значення $r_3 = 0$, то характер кривих визначається виключно величиною r_4. Область значень $r_4 < 3$ розподіляється на три підобласті, для кожної з яких криві розподілу набувають певного вигляду.

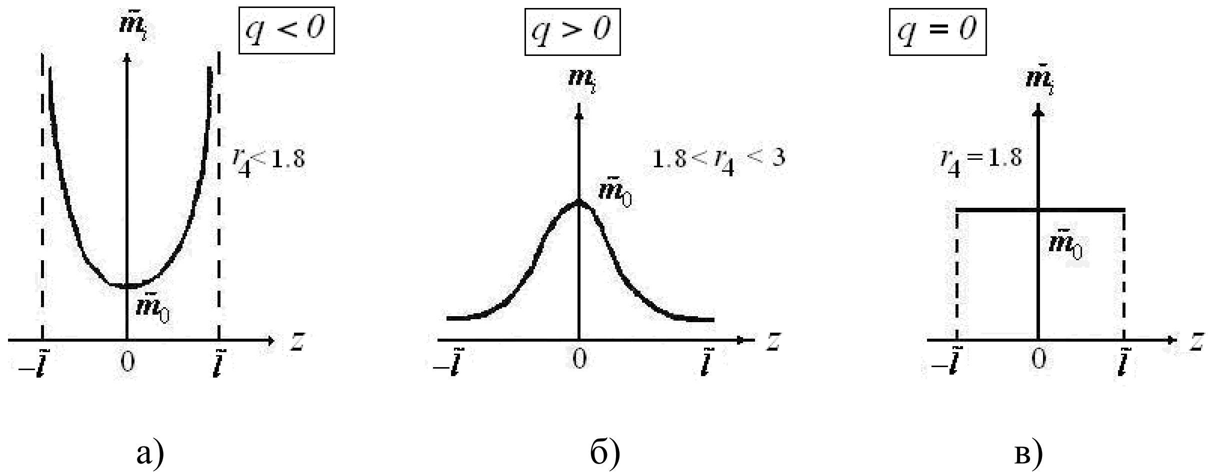


Рис. 4.6 – Криві розподілу другого типу Пірсона [1]

Якщо $r_4 < 1.8$, то $q < 0$ і маємо симетричний $\cup$ – подібний розподіл (рис. 4.6 а).

За умови $1.8 < r_4 < 3$, $q > 0$, що відповідає дзвонеподібному розподілу (рис. 4.6 б), крива розподілу Пірсона II типу суттєво відрізняється від кривої нормального розподілу. По-перше, область існування другого типу Пірсона є обмеженою: $-\tilde{l} \le z \le \tilde{l}$, а нормального розподілу – необмеженою: $-\infty < x < \infty$. По-друге, вона є більш стислою $(E < 0)$, ніж крива нормального розподілу.

Нарешті, за умови $r_4 = 1.8$, $q = 0$ та $\tilde{m}_i = \tilde{m}_0$ для всіх значень z_i отримаємо рівномірний розподіл (рис. 4.6 в).

Види кривих розподілів Пірсона II типу, що зображені на рис. 4.6, необхідно враховувати при апроксимації емпіричних розподілів випадкових величин другим типом розподілів Пірсона.

Алгоритм апроксимації емпіричного розподілу другим типом розподілів Пірсона складається з 4-х основних етапів.

Відомо, що при дослідженнях законів розподілу випадкових величин, *перш за все,* необхідно від простої статистичної сукупності $X : x_1, x_2, x_3, \dots, x_i, \dots, x_n$ перейти до згрупованого ряду $(\tilde{x}_i ; m_i)$ або $(\tilde{x}_i ; \hat{p}_i)$. Статистично оцінити перший початковий $(\hat{\nu}_1 = \overline{x})$, другий центральний $(\hat{\mu}_2)$ моменти розподілу, отримати незсунену, ефективну і умотивовану оцінку дисперсії $\left(S_x^2\right)$ та середнього квадратичного відхилу (S_x) випадкової величини, що досліджується. Розрахувати коефіцієнти асиметрії $(As = \hat{r}_3)$ та ексцесу (E). Графічно представити отриманий емпіричний розподіл. Оцінити статистики S, æ і, якщо для æ виконується рівність æ = 0 та умови $r_3 \approx 0$, $r_4 < 3$, формулюють гіпотезу на заданому рівні значущості (розділ 3) про можливість апроксимації емпіричного розділу II типом розподілів Пірсона.

На *другому етапі* за відповідними формулами оцінюють параметри другого типу розподілів Пірсона, а саме: q, $\tilde{l}$, $\tilde{m}_0$ та необхідні статистики.

На *третьому етапі* розраховують інтервальні теоретичні частоти $\tilde{m}_i$ за допомогою основного рівняння II типу (4.87), перейшовши спочатку від розмірних значень середин часткових інтервалів $\tilde{x}_i$ до безрозмірних величин z_i $\left(i = \overline{1,k}\right)$. Використовуючи рівняння (4.75) та умови $r_3 = 0$, $\hat{x} = \overline{x}$, маємо :

$$z_i = \frac{\tilde{x}_i - \overline{x}}{c}. \tag{4.91}$$

Розраховані для кожної градації інтервальні теоретичні частоти $\tilde{m}_i$ необхідно порівняти з відповідними емпіричними частотами m_i і на основі отриманих характеристик (вираз (4.92)) побудувати полігони цих розподілів та з'ясувати спочатку якісні розбіжності між m_i та $\tilde{m}_i$:

$$\begin{array}{l} \tilde{x}_1,\ \ \tilde{x}_2,\ \ldots,\ \ \tilde{x}_i,\ \ldots,\ \tilde{x}_k \\ m_1,\ \ m_2,\ \ldots,\ \ m_i,\ \ldots,\ m_k\ . \\ \tilde{m}_1,\ \ \tilde{m}_2,\ \ldots,\ \ \tilde{m}_i,\ \ldots,\ \tilde{m}_k \end{array} \qquad (4.92)$$

Метод визначення кількісних розбіжностей між інтервальними емпіричними та розрахованими теоретичними частотами II типу Пірсона за допомогою відомих критеріїв (див. підрозділ 4.3) дає змогу отримати остаточну відповідь щодо можливості (з заданою ймовірністю) апроксимації емпіричного розподілу другим типом розподілів Пірсона. Тому на *четвертому етапі* треба перевірити статистичну гіпотезу про відповідність емпіричного та теоретичного розподілів. Оскільки перевірка цієї гіпотези відносно типу розподілу є самостійною задачею, алгоритм її розв'язання розглядався у підрозділі 4.3.

4.4.5 Третій тип розподілів Пірсона

У відповідності до загальної теорії розподілів Пірсона (пункт 4.4.1), третій тип *спостерігається при будь-яких значеннях статистики капа* $(-\infty < æ < \infty)$, яка розраховується за формулою (4.72).

Основне рівняння III типу розподілів Пірсона має вигляд:

$$\tilde{m}_i = \tilde{m}_0 \left(1 + \frac{z_i}{l'}\right)^p e^{-pz_i/l'}, \qquad (4.93)$$

де $p,\ l',\ \tilde{m}_0$ – параметри цього розподілу;

z_i – змінна, яка розраховуються так:

$$z_i = \frac{\tilde{x}_i - \hat{x}}{c}, \qquad (4.94)$$

де

$$\hat{x} = \overline{x} - \frac{r_3 \sigma_x}{2}. \qquad (4.95)$$

Параметр форми p визначається рівнянням (4.96):

$$p = \frac{4}{r_3^2} - 1. \qquad (4.96)$$

Параметр масштабу l' розраховується за формулою (4.97):

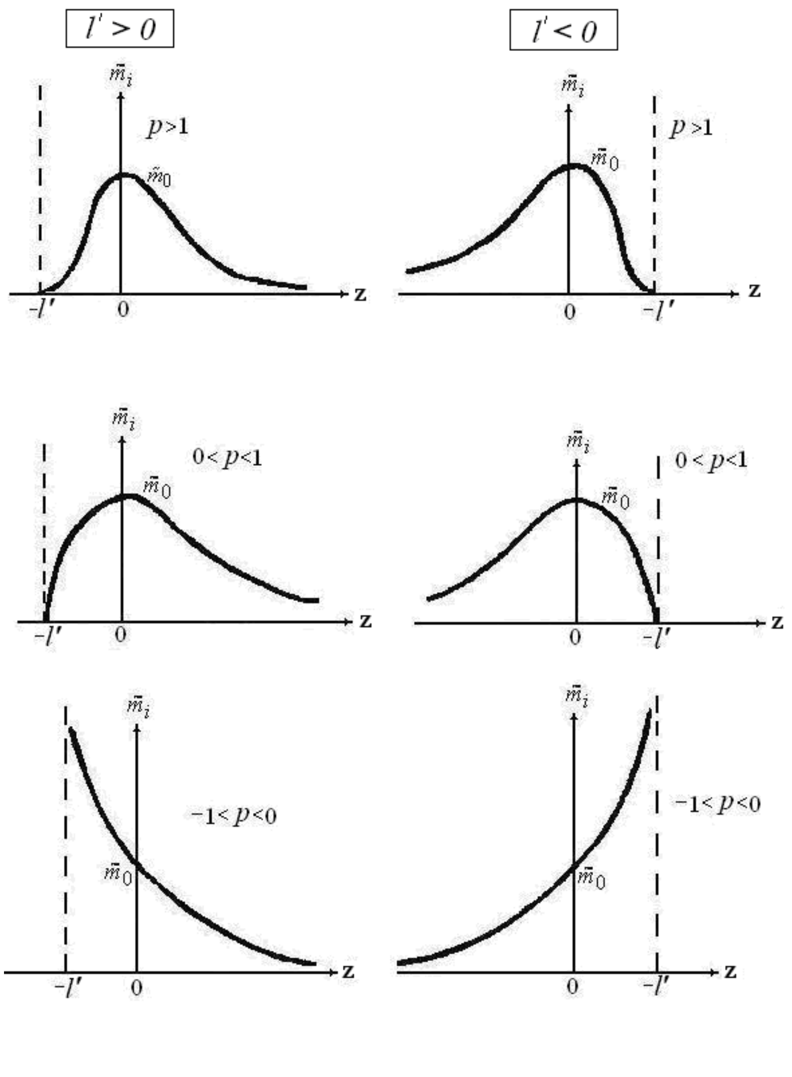
$$l' = \sigma\left(\frac{2}{r_3} - \frac{r_3}{2}\right). \qquad (4.97)$$

Параметр $\tilde{m}_0$ *(теоретична частота за умови* $z_i = 0$) можна отримати за допомогою рівняння (4.98):

$$\tilde{m}_0 = \frac{n}{|l'|} \frac{p^{p+1}}{e^p \Gamma(p+1)} \;, \qquad (4.98)$$

в якому $\Gamma(x)$ – гамма функція відповідного аргументу. Її значення для заданого x наводяться у таблицях (Додаток Е).

Як випливає з рівняння (4.93), *III тип розподілів Пірсона об'єднує 6 різновидів кривих* (рис. 4.7). Криві, що відносяться до цього типу розподілів Пірсона, можна об'єднати у дві групи.



а) б)

Рис. 4.7 – Криві розподілу третього типу Пірсона [1]

1) $l' > 0$. За умови додатних значень параметра масштабу l' спостерігається *три* різновиди кривих розподілу. Загальним для них є дві властивості: за умови $z_i \to \infty$, $\tilde{m}_i \to 0$ (експоненціальна функція спадає швидше, ніж зростає степенева функція); за умови $z_i = -l'$, $\tilde{m}_i = 0$, тобто точка $z_i = -l'$ визначає початок розподілу.

2) $l' < 0$. Від'ємним значенням параметра масштабу теж відповідає *три* різновиди кривих розподілу, які відрізняються від попередніх кривих протилежною асиметрією.

За умови $l' > 0$ розподіли мають правосторонню асиметрію (рис. 4.7 а), а за умови $l' < 0$ – лівосторонню асиметрію (рис. 4.7 б).

Таким чином, III тип розподілів Пірсона в залежності від параметра масштабу l' включає 6 видів кривих, які необхідно врахувати при апроксимації емпіричних розподілів III типом розподілів Пірсона.

Алгоритм апроксимації емпіричного розподілу III типом розподілів Пірсона складається, як і для інших теоретичних розподілів, із 4-х основних етапів.

Перш за все необхідно від простої статистичної сукупності $X: x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ (де n – об'єм вибірки) перейти до згрупованого ряду вигляду $(\tilde{x}_i; m_i)$ або $(\tilde{x}_i; \hat{p}_i)$ $\forall\, i = \overline{1,k}$. Статистично оцінити перший початковий $(\hat{\nu}_1 = \bar{x})$ та другий центральний $(\hat{\mu}_2)$ моменти розподілу; отримати незсунені, ефективні та умотивовані оцінки дисперсії $\left(S_x^2\right)$ і середнього квадратичного відхилення (S_x). Розрахувати коефіцієнти асиметрії $(As = \hat{r}_3)$ та ексцесу (E). Графічно представити отриманий емпіричний розподіл частот. Оцінити статистики S, æ і, якщо з заданою ймовірністю немає можливості апроксимувати I чи II типом розподілів Пірсона (чи нормальним розподілом), то треба зробити спробу апроксимувати емпіричний розподіл III-м типом розподілів Пірсона.

На *другому етапі* оцінюють параметри теоретичного розподілу, а саме q, l', $\tilde{m}_0$ за формулами (4.96)–(4.98) та необхідні для цього статистики.

На *третьому етапі* розраховують інтервальні теоретичні частоти $\tilde{m}_i$ за допомогою основного рівняння III типу розподілів Пірсона (4.93). Але спочатку треба перейти від розмірних значень середин часткових інтервалів $\tilde{x}_i \left(i = \overline{1,k}\right)$ до безрозмірних величин $z_i \left(i = \overline{1,k}\right)$, використовуючи формули (4.94) та (4.95).

Розраховані для кожної градації $\left(i = \overline{1,k}\right)$ інтервальні теоретичні частоти $\tilde{m}_i$ треба порівняти з відповідними емпіричними частотами m_i і на основі отриманих характеристик (вираз (4.99)) побудувати полігони емпіричних та теоретичних частот:

$$\begin{matrix} \tilde{x}_1, & \tilde{x}_2, & \dots, & \tilde{x}_i, & \dots, & \tilde{x}_k \\ m_1, & m_2, & \dots, & m_i, & \dots, & m_k \\ \tilde{m}_1, & \tilde{m}_2, & \dots, & \tilde{m}_i, & \dots, & \tilde{m}_k \end{matrix} . \tag{4.99}$$

Побудова полігонів цих розподілів допоможе з'ясувати наявність чи відсутність якісних розбіжностей між m_i та $\tilde{m}_i$.

Метод визначення кількісних розбіжностей між інтервальними емпіричними та розрахованими на основі рівняння (4.93) теоретичними частотами III типу розподілів Пірсона за допомогою відомих критеріїв (див. підрозділ 4.3) допоможе отримати остаточну відповідь про можливість (з заданою ймовірністю) апроксимації емпіричного розподілу III типом розподілів Пірсона. Тому дослідження завершується *на четвертому етапі* перевіркою статистичної гіпотези про відповідність емпіричного та підібраного теоретичного розподілів, тобто з'ясовується статистично значущі чи випадкові розбіжності між m_i та $\tilde{m}_i$. Оскільки перевірка статистичних гіпотез відносно типу розподілу є самостійною задачею, алгоритм її розв'язання розглядався окремо у підрозділі 4.3.

4.5 Розподіл Пуассона

У багатьох випадках, коли досліджують явища, що рідко реалізуються у природі, ці події розглядаються як випадкові і можуть підпорядковуватися розподілу Пуассона. У метеорологічній практиці такими величинами є кількість днів з грозою, хуртовиною, ожеледдю тощо. *Ці дискретні випадкові величини (X) можуть набувати тільки значення з множини цілих натуральних чисел 0, 1, 2, ...*

За законом Пуассона ймовірність того, що $X = m$, визначається формулою:

$$P(X = m) = \frac{\lambda^m}{m!} e^{-\lambda} \qquad (m = 0,1,2,...). \tag{4.100}$$

де $\lambda > 0$ – *параметр даного розподілу.*

Визначимо математичне сподівання та дисперсію такої випадкової величини.

$$\begin{aligned} m_x &= \sum_{m=0}^{\infty} mP(x = m) = \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} \end{aligned} . \tag{4.101}$$

Як відомо, для нескінченного ряду

$$\sum_{m=0}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}=e^{\lambda}. \qquad (4.102)$$

Отже,

$$m_x=e^{-\lambda}\lambda e^{\lambda}=\lambda\,. \qquad (4.103)$$

Стала величина λ є математичним сподіванням випадкової величини, яка має розподіл Пуассона. Це означає, що розподіл Пуассона однозначно визначається математичним сподіванням випадкової величини. Дисперсію цієї величини отримаємо за допомогою відомого співвідношення

$$\sigma_x^2=\nu_2-\nu_1^2\,. \qquad (4.104)$$

Знайдемо спочатку другий початковий момент

$$\begin{aligned}\nu_2 &= \sum_{m=0}^{\infty}m^2\frac{\lambda^m}{m!}e^{-\lambda}=\lambda\sum_{m=0}^{\infty}m\frac{\lambda^{m-1}}{(m-1)!}e^{-\lambda}=\\ &\lambda\sum_{m=1}^{\infty}[(m-1)+1]\frac{\lambda^{m-1}}{(m-1)!}e^{-\lambda}=\\ &=\lambda\left[\sum_{m=1}^{\infty}(m-1)\frac{\lambda^{m-1}}{(m-1)!}e^{-\lambda}+\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}\right]=\\ &=\lambda\left[\lambda e^{-\lambda}\sum_{m=2}^{\infty}\frac{\lambda^{m-2}}{(m-2)!}+1\right]=\\ &=\lambda(\lambda+1)\,.\end{aligned} \qquad (4.105)$$

Оскільки $\nu_1=m_x=\lambda$, то

$$\sigma_x^2=\lambda^2+\lambda-\lambda^2=\lambda\,. \qquad (4.106)$$

Таким чином,

$$\sigma_x^2=m_x=\lambda\ . \qquad (4.107)$$

Отже, дисперсія випадкової величини, що має розподіл Пуассона, чисельно дорівнює її математичному сподіванню.

Третій та четвертий центральні моменти розподілу Пуассона визначаються формулами:

$$\mu_3 = \lambda \ , \tag{4.108}$$

$$\mu_4 = \lambda + 3\lambda^2 . \tag{4.109}$$

Тому для коефіцієнтів асиметрії та ексцесу можна отримати такі формули:

$$A_s = r_3 = \frac{\mu_3}{\sigma_x^3} = \frac{\lambda}{(\sqrt{\lambda})^3} = \frac{1}{\sqrt{\lambda}}, \tag{4.110}$$

$$E = \frac{\mu_4}{\sigma_x^4} - 3 = \frac{\lambda + 3\lambda^2}{\lambda^2} - 3 = \frac{1}{\lambda}. \tag{4.111}$$

Таким чином, основна властивість закону Пуассона визначається рівнянням:

$$\lambda = m_x = \sigma_x^2 . \tag{4.112}$$

Очевидно, для статистичної оцінки цього параметра повинні виконуватися співвідношення:

$$\hat{\lambda} = \hat{m}_x = \hat{\sigma}_x^2 . \tag{4.113}$$

А це означає, що

$$\hat{\lambda} = \overline{x} \approx S_x^2 . \tag{4.114}$$

Після визначення оцінки параметра λ на основі згрупованого ряду по таблицях знаходять імовірності $P(X = m)$ за умови $m = 0, 1, 2, 3\ldots$ (Додаток Ж). Для розрахунків інтервальних теоретичних частот використовується таке співвідношення:

$$\tilde{m}_i = n\,P(X = m). \tag{4.115}$$

Як і у попередніх випадках, дослідження закінчується порівнянням емпіричних (m_i) та отриманих за формулою (4.115) теоретичних $(\tilde{m}_i)$ частот. Якісне порівняння частот проводиться на основі полігонів, а кількісне – з використанням відомих критеріїв. Тому далі треба перевірити статистичну гіпотезу про відповідність на заданому рівні значущості емпіричного та теоретичного розподілів. Оскільки перевірка цієї гіпотези відносно типу розподілу є самостійною задачею, алгоритм її розв’язання розглядався у підрозділі 4.3.

Запитання для самоперевірки

1. Що називається «законом розподілу» та які з них найчастіше використовуються при статистичних дослідженнях властивостей випадкових величин?
2. З яких етапів складається дослідження закону розподілу випадкової величини?
3. Якими функціями можна представити теоретичний розподіл? Дати їм визначення та перелічити основні властивості.
4. Чому закон розподілу можливо представляти інтервальними теоретичними частотами?
5. Чим відрізняються прості статистичні сукупності від згрупованих?
6. Як розраховують межі часткових інтервалів, їх кількість та довжину?
7. Який сенс гістограми та полігону розподілу випадкової величини?
8. Перелічити основні властивості нормального розподілу.
9. Який параметр для кривої нормального розподілу є параметром форми? параметром масштабу?
10. Якщо випадкова величина нормально розподілена, з якою ймовірністю вона буде знаходитися в інтервалі $[m_x \pm \sigma_x]$? $[m_x \pm 2\sigma_x]$? $[m_x \pm 3\sigma_x]$?
11. Підкреслити основні умови, за яких дослідник може сформулювати гіпотезу про можливість апроксимації емпіричного розподілу нормальним законом.
12. Який зв'язок між випадковою величиною X та нормованою величиною t у випадку нормального розподілу?
13. Як розрахувати інтервальні теоретичні частоти нормального розподілу, використовуючи функцію $f(t)$? Які властивості вона має?
14. Як розрахувати інтервальні теоретичні частоти нормального розподілу, використовуючи інтеграл імовірностей $\Phi(t)$? Які властивості ця функція має?
15. Від яких статистичних оцінок основних моментів розподілу залежить статистика æ?
16. Підкреслити основні умови, за яких дослідник може сформулювати гіпотезу про можливість апроксимації емпіричного розподілу І типом розподілів Пірсона.
17. Які параметри для кривої І типу розподілів Пірсона є параметрами форми? параметрами масштабу?
18. Який зв'язок між випадковою величиною x та безрозмірною величиною z?
19. За якою формулою розраховуються інтервальні теоретичні частоти І типу розподілів Пірсона?

20. Підкреслити основні умови за яких дослідник може сформулювати гіпотезу про можливість апроксимації емпіричного розподілу II типом розподілів Пірсона.
21. Які параметри для кривої II типу розподілів Пірсона є параметрами форми? параметрами масштабу?
22. За якою формулою розраховуються інтервальні теоретичні частоти II типу розподілів Пірсона?
23. Чим відрізняється крива нормального розподілу від кривої II типу розподілів Пірсона?
24. Підкреслити основні умови за яких дослідник може сформулювати гіпотезу про можливість апроксимації емпіричного розподілу III типом розподілів Пірсона.
25. Які параметри для кривої III типу розподілів Пірсона є параметрами форми? параметрами масштабу?
26. За якою формулою розраховуються інтервальні теоретичні частоти III типу розподілів Пірсона?
27. Як розраховуються часткові інтервали, їх кількість, інтервальні емпіричні та теоретичні частоти при апроксимації емпіричного розподілу розподілом Пуассона?
28. Якщо випадкова величина підпорядковується розподілу Пуассона, яке співвідношення між першим початковим та другим центральним моментами розподілу?
29. За допомогою якого рівняння розраховуються інтервальні теоретичні частоти розподілу Пуассона?
30. В якому випадку використовується розподіл Пуассона при апроксимації емпіричного розподілу?

5. КОРЕЛЯЦІЙНИЙ ЗВ’ЯЗОК МІЖ ДВОМА ВИПАДКОВИМИ ВЕЛИЧИНАМИ

5.1 Форма та тіснота кореляційного зв'язку

5.1.1 Поняття про статистичні зв'язки між двома випадковими величинами

Дослідник має всі підстави розглядати гідрометеорологічні величини як випадкові величини, незважаючи на те, що їх зміни обумовлені певними фізичними закономірностями. Така їх властивість пов'язана з тим, що зв'язки між випадковими величинами спричиняються взаємодією багатьох факторів, і у нас немає, як правило, достатніх знань про те, який саме зв'язок, або зв'язки, обумовили змінювання цієї величини у термін її вимірювання. Більше того, у деяких випадках ще є відсутніми повні уявлення про причинно-наслідкові залежності у природних явищах, що спричиняють гідрометеорологічні процеси. Деяка міра невизначеності має місце, коли йдеться про майбутній стан атмосфери чи об'єктів гідросфери, тобто про гідрометеорологічне прогнозування.

У природі досить часто виникають ситуації, коли між деякими двома випадковими величинами проявляються статистичні зв'язки. Вони можуть бути *функціональними* або *стохастичними*.

Функціональною залежністю між двома випадковими величинами називається така залежність, коли зі зміною однієї випадкової величини змінюється інша. І можливому значенню однієї випадкової величини відповідає тільки одне значення іншої.

Стохастичним називають такий *зв'язок* між випадковими величинами, коли зміна однієї з них приводить до зміни закону розподілу іншої. Стохастичний зв'язок між двома випадковими величинами спостерігається, наприклад, коли існують загальні випадкові фактори, впливаючи як на одну, так і на другу величину поряд з іншими неоднаковими для обох величин випадковими факторами. Наприклад, якщо Y є деяка функція від випадкових величин $z_1,\ z_2, \dots,\ z_m,\ \upsilon_1,\ \upsilon_2, \dots,\ \upsilon_к$:

$$Y = f\ (z_1,\ z_2, \dots,\ z_m,\ \upsilon_1,\ \upsilon_2, \dots,\ \upsilon_к),$$

а X – функція від тих же випадкових величин $z_1,\ z_2, \dots,\ z_m$ і деякої сукупності інших випадкових величин $u_1,\ u_2, \dots,\ u_к$:

$$X = \varphi\left(z_1,\ z_2, \dots,\ z_m,\ u_1,\ u_2, \dots,\ u_к\right),$$

то залежність між випадковими величинами X та Y буде стохастичною.

Найбільш важливі особливості стохастичного зв'язку виявляються у тих змінюваннях, які зазнає центр умовного розподілу однієї величини при змінюванні іншої. Умовний розподіл – це такий розподіл між двома випадковими величинами, коли на одну з них накладається певна умова.

Якщо припустити, що *умовний розподіл* є нормальним, з щільністю ймовірності

$$f\left(y/x\right)=\frac{1}{\sigma_{y/x}\sqrt{2\pi}}\exp\left[-\frac{\left(y-m_{y/x}\right)^2}{2\sigma^2_{y/x}}\right], \tag{5.1}$$

де

$m_{y/x}=M\left[Y/X=x\right]$ – *умовне математичне сподівання випадкової величини Y, коли на випадкову величину X накладається певна умова X=x;*

$\sigma^2_{y/x}=M\left\{\left[y-m_{y/x}\right]^2\right\}$ – *умовна дисперсія випадкової величини Y.*

Центром цього розподілу є умовне математичне сподівання.

Отже, *якщо при змінюванні однієї з випадкових величин змінюється умовне математичне сподівання іншої, то такий зв'язок між цими випадковими величинами називається кореляційним.*

Кореляційну залежність можна трактувати як функціональну залежність умовного математичного сподівання однієї випадкової величини від значення іншої:

$$m_{y/x}=f(x). \tag{5.2}$$

Функцію $f(x)$ називають *функцією регресії* величини Y на (по) X. Рівняння (5.2) називається *рівнянням регресії*. Можна розглядати і кореляційний зв'язок між величинами X та Y, а саме:

$$m_{x/y}=\varphi(y). \tag{5.3}$$

Треба мати на увазі, що на основі вибірок випадкових величин

$$X:\ x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_n$$

та (5.4)

$$Y:\ y_1,\ y_2,\ \ldots,\ y_i,\ \ldots,\ y_n$$

ми можемо отримати лише *оцінки* відповідних *умовних математичних сподівань*, якими є *умовні середні значення*:

$$\hat{m}_{y/x}=\bar{Y}/_{X=x}=\bar{y}(x);\quad \hat{m}_{x/y}=\bar{X}/_{Y=y}=\bar{x}(y). \tag{5.5}$$

Їм відповідають *оціночні рівняння регресії*:

$$\bar{y}(x)=\hat{f}(x); \tag{5.6}$$

$$\bar{x}(y)=\hat{\varphi}(y). \tag{5.7}$$

Таким чином, задача дослідника полягає у тому, щоб отримати рівняння регресії, яке б із заданою ймовірністю описувало кореляційну залежність між випадковими величинами, що розглядаються. І починати дослідження треба з побудови кореляційного графіка, який дозволяє якісно визначити форму та тісноту статистичного зв’язку між випадковими величинами.

5.1.2 Якісне уявлення про тісноту та форму кореляційного зв’язку між двома випадковими величинами

Кореляційні зв’язки між випадковими величинами *характеризуються формою* та *тіснотою*. Ці ознаки статистичного зв’язку можна визначати як якісно (за допомогою кореляційного графіка), так і кількісно (за допомогою відповідних коефіцієнтів).

Що стосується *форми кореляційного зв’язку* між двома випадковими величинами, то дуже часто реалізуються зв’язки, що описуються:

– *лінійним рівнянням регресії*:

$$\bar{y}(x) = ax + b, \tag{5.8}$$

де a і b – коефіцієнти цього рівняння.

Лінійний кореляційний зв’язок може бути *прямим* і *оберненим*. Якщо при збільшенні (зменшенні) однієї випадкової величини відбувається збільшення (зменшення) умовного математичного сподівання (умовного середнього) іншої, то такий *зв’язок* називають *прямим*. Коли збільшенню (зменшенню) однієї випадкової величини відповідає зменшення (збільшення) умовного математичного сподівання (умовного середнього) іншої, то це свідчить про *обернений лінійний кореляційний зв’язок* між цими випадковими величинами.

Крім лінійної кореляційної залежності (*прямої* чи *оберненої*) зустрічаються *нелінійні кореляційні зв’язки*, які відбиваються відповідними рівняннями регресії:

– *параболічним*:

$$\bar{y}(x) = a_0 + a_1 x + a_2 x^2; \tag{5.9}$$

– *показниковим*:

$$\bar{y}(x) = ab^{cx}, \tag{5.10}$$

частинним випадком якого є *експоненціальне рівняння регресії*:

$$\bar{y}(x) = ae^{bx}; \tag{5.11}$$

– *гіперболічним*:

$$\bar{y}(x) = \frac{a}{x^b}. \qquad (5.12)$$

Якісне уявлення про тісноту та форму кореляційного зв'язку між величинами X та Y можна отримати, побудувавши *кореляційний графік (діаграму розсіювання точок або поле кореляції)* на площині XOY в координатах $(x_i; y_i)$. Якщо точки тісно групуються біля деякої осередненої лінії, то це свідчить про те, що кореляційний *зв'язок є тісним*. Чим більшим є розкид точок на графіку, тим слабкішим є кореляційний зв'язок.

Приклад тісного прямого лінійного кореляційного зв'язку наводиться на рис. 5.1.

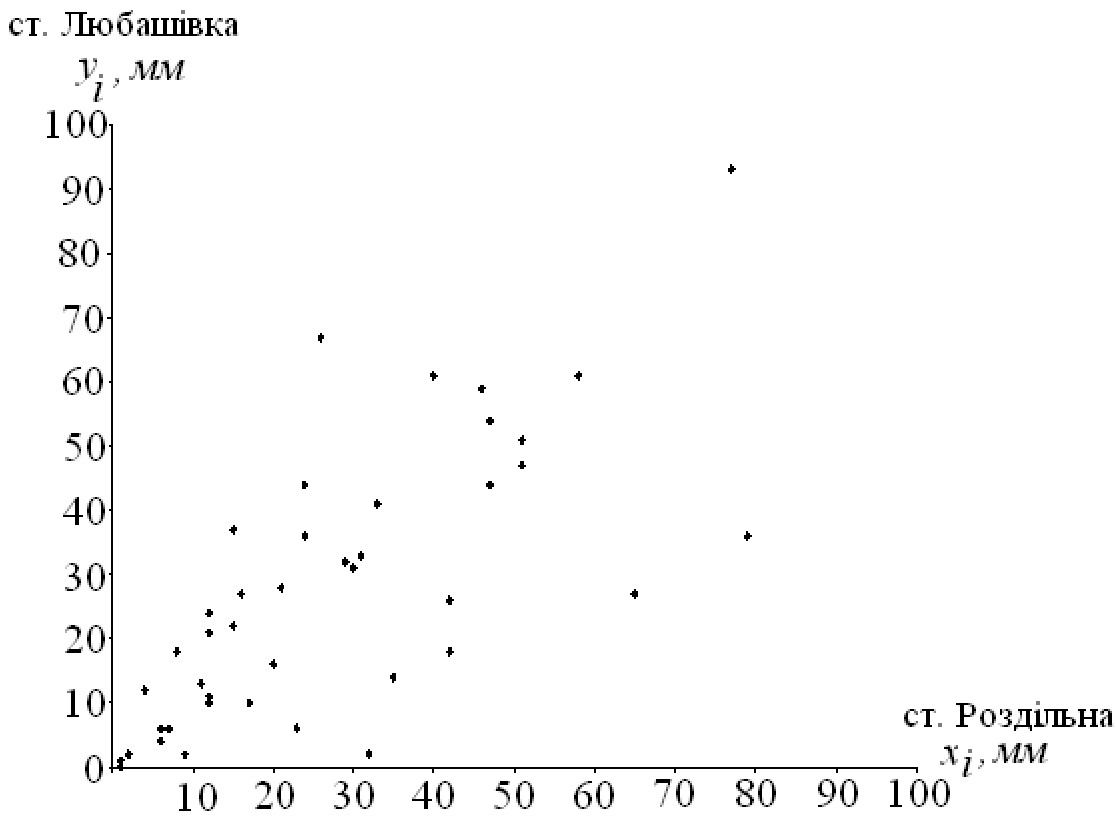


Рис. 5.1 – Поле кореляції місячної кількості опадів (жовтень) [1]

5.1.3 Кількісна міра тісноти кореляційного зв'язку між двома випадковими величинами

У статистичних дослідженнях якісне визначення тісноти та форми кореляційного зв'язку між двома випадковими величинами дає можливість припустити наявність (чи відсутність) будь-якої форми кореляційного зв'язку. Щоб мати уявлення про *кількісну міру тісноти кореляційного зв'язку використовують кореляційне відношення*.

Якщо розглядати у загальному плані змінювання випадкової величини Y, то на величину Y крім величини X можуть діяти й інші випадкові величини. Тому дисперсія випадкової величини Y, яку називають *повною дисперсією* (σ_y^2), складається з двох частин:

$$\sigma_y^2 = \sigma_{y/x}^2 + \delta_{y/x}^2 . \tag{5.13}$$

Член

$$\sigma_{y/x}^2 = M\left\{\left[y - \bar{y}(x)\right]^2\right\} \tag{5.14}$$

характеризує розсіювання точок на кореляційному графіку відносно лінії регресії, тобто відбиває вплив інших діючих факторів на випадкову величину Y.

Другий член рівняння (5.13)

$$\delta_{y/x}^2 = M\left\{\left[\bar{y}(x) - m_{y/x}\right]^2\right\} \tag{5.15}$$

є мірою розсіювання вибіркових ліній регресії $\bar{y}(x) = \hat{f}(x)$ відносно генеральної лінії регресії $m_{y/x} = f(x)$ і характеризує саме вплив випадкової величини X на випадкову величину Y.

Параметр

$$\eta_{y/x}^2 = \frac{\delta_{y/x}^2}{\sigma_y^2} \tag{5.16}$$

є показником тісноти кореляційного зв'язку між випадковими величинами Y та X. Він носить назву *кореляційного відношення*. Всі можливі значення кореляційного відношення визначаються виразом:

$$0 \le \eta_{y/x}^2 \le 1. \tag{5.17}$$

Чим ближчим є значення кореляційного відношення до одиниці, тим тіснішим є кореляційний зв'язок між випадковими величинами X та Y, і навпаки.

Статистична оцінка кореляційного відношення знаходиться на основі оцінок розглянутих вище складових повної дисперсії:

$$S_{y/x}^2 = \hat{\sigma}_{y/x}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}(x_i)\right]^2; \quad (5.18)$$

$$\hat{\delta}_{y/x}^2 = \frac{1}{k}\sum_{j=1}^{k}\left[\bar{y}_j(x) - \bar{y}\right]^2; \quad (5.19)$$

$$\hat{\eta}_{y/x}^2 = \frac{\hat{\sigma}_{y/x}^2}{S_y^2}, \quad (5.20)$$

де

$$S_y^2 = S_{y/x}^2 + \hat{\delta}_{y/x}^2 . \quad (5.21)$$

5.1.4 Коефіцієнт кореляції як кількісна міра лінійного кореляційного зв'язку між двома випадковими величинами

Якщо кореляційний зв'язок між величинами Y та X є лінійним, то кореляційне відношення вироджується у *параметр* r_{xy}^2, який називається *коефіцієнтом детермінації*.

Параметр

$$r_{xy} = \sqrt{r_{xy}^2} \quad (5.22)$$

називається *коефіцієнтом кореляції* і розраховується на основі статистичних сукупностей (5.4) за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y}. \quad (5.23)$$

Він може приймати значення з множини

$$-1 \le r_{xy} \le 1. \quad (5.24)$$

Коефіцієнт кореляції характеризує степінь наближення кореляційного зв'язку між випадковими величинами X та Y до лінійної функціональної залежності.

Додатні значення коефіцієнта кореляції $(r_{xy} > 0)$ позначають наявність *прямого лінійного зв'язку* між випадковими величинами X та Y. *Обернений лінійний кореляційний зв'язок* характеризується *від'ємними* значеннями коефіцієнта кореляції $(r_{xy} < 0)$.

Чим більшим за модулем є коефіцієнт кореляції $|r_{xy}|$, *тим тіснішим буде лінійний кореляційний зв'язок* між випадковими величинами, і навпаки. Якщо $r_{xy} = 0$, то це означає, що *лінійний кореляційний зв'язок між величинами* X *та* Y *відсутній*, але це не означає, що відсутнім є кореляційний зв'язок іншої форми.

5.2 Перевірка гіпотези про статистичну значущість коефіцієнта кореляції

Після отримання точкової статистичної оцінки коефіцієнта кореляції r_{xy} необхідно оцінити вірогідність лінійного кореляційного зв'язку між випадковими величинами X та Y.

І перш ніж приступати до побудови лінійного рівняння регресії треба перевірити гіпотезу про статистичну значущість коефіцієнта кореляції.

Перевірку цієї гіпотези необхідно провести після того, як на основі вибірок випадкових величин X та Y отримана статистична оцінка коефіцієнта кореляції визиває сумнів. У такому разі треба удосконалитися в тому, що статистична оцінка $r_{xy} = \hat{\rho}_{xy}$ не є випадковою.

Висновок про це можна зробити тільки після перевірки гіпотези про статистичну значущість оцінки коефіцієнта кореляції, яка, як відомо, визначається на основі випадкових вибірок. Нульову гіпотезу формулюють так:

H_0: На рівні значущості α коефіцієнт кореляції є статистично Незначущим, як і його оцінка:

$$\rho_{xy} = 0; \qquad \hat{\rho}_{xy} = r_{xy} = 0.$$

Альтернативна гіпотеза H_1, очевидно, є такою:

H_1: На рівні значущості α коефіцієнт кореляції є статистично Значущим, як і його оцінка:

$$\rho_{xy} \neq 0; \qquad \hat{\rho}_{xy} = r_{xy} \neq 0.$$

Перевірка гіпотези H_0 відносно H_1 втілюється за допомогою критерію Стьюдента. У загальному випадку, за умови *малих об'ємів* вибірок $(n < 50)$, критерій Стьюдента визначається таким чином:

$$t = \frac{|\hat{z}|}{\sigma_z}, \tag{5.25}$$

де

$$\hat{z} = \frac{1}{2}\ln\frac{1+r_{xy}}{1-r_{xy}}, \tag{5.26}$$

$$\sigma_z = \frac{1}{\sqrt{n-3}}. \tag{5.27}$$

Якщо *ряди мають представництво* $(n > 50)$, для побудови критерію Стьюдента можна використовувати безпосередньо отриману оцінку коефіцієнта кореляції r_{xy}.

Тоді:

$$t = \frac{|r_{xy}|}{\sigma_r}, \tag{5.28}$$

де

$$\sigma_r = \frac{1-r_{xy}^2}{\sqrt{n-1}}. \tag{5.29}$$

Розраховане за формулою (5.25) або (5.28) (в залежності від об'єму вибірки) значення критерію Стьюдента порівнюють з $t_{кр}(\alpha,\nu)$, яке наводиться в таблицях (Додаток А). Рівень значущості α задає дослідник та розраховує число степенів вільності за формулою $\nu = n-1$.

Якщо

$$t < t_{кр}(\alpha,\nu), \tag{5.30}$$

то приймається гіпотеза H_0 про статистичну незначущість, тобто про випадковість отриманої статистичної оцінки коефіцієнта кореляції.

У протилежному випадку, якщо

$$t > t_{кр}(\alpha,\nu), \tag{5.31}$$

гіпотеза H_0 відкидається й береться альтернативна гіпотеза H_1 про те, що коефіцієнт кореляції є статистично значущим. А це означає, що він дійсно виражає характер лінійного кореляційного зв'язку між випадковими величинами X та Y і є підстави виразити цей статистичний зв'язок генеральним лінійним рівнянням регресії:

$$m_{y/x} = Ax + B, \tag{5.32}$$

де

$m_{y/x}$ – умовне математичне сподівання випадкової величини Y;

A і B – коефіцієнти генерального рівняння лінійної регресії.

Статистичною оцінкою генерального рівняння регресії є рівняння:

$$\overline{y}(x) = ax + b,$$

де

$\overline{y}(x)$ – умовне середнє значення випадкової величини Y;

$\hat{A} = a$ і $\hat{B} = b$.

5.3 Метод найменших квадратів (МНК)

Якщо статистична оцінка коефіцієнта кореляції не є випадковою, приступають до побудови *лінійної регресійної моделі* вигляду (5.32) на основі статистичних сукупностей (5.4). А це означає, що треба знайти *оцінку рівняння* (5.32), тобто $\overline{y}(x) = ax + b$. Причому, коефіцієнти регресії a і b є статистичними оцінками коефіцієнтів *генерального рівняння регресії*: $\hat{A} = a$ і $\hat{B} = b$.

Зрозуміло, що отримана модель взаємозв’язку між випадковими величинами X та Y повинна бути адекватною тому процесу, який моделюється.

Можуть бути різні степені адекватності прогностичних моделей. Тому при моделюванні треба визначити кількісну міру адекватності. Таку *кількісну міру називають критерієм якості або функцією цілі.*

Критерій якості вибирають у залежності від характеру задачі, яка вирішується. При побудові статистичних моделей найчастіше використовують такий критерій:

$$\Delta^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2 = min, \qquad (5.33)$$

де

y_i – ординати точок кореляційного поля;

$f(x_i)$ – ординати точок на лінії регресії при значенні незалежної змінної x_i.

Сенс критерію Δ^2 полягає в тому, що для опису взаємозв'язку між випадковими величинами X та Y вибирають такі параметри рівняння регресії $\overline{y}(x)$, які дають мінімум суми квадратів різниць між ординатами експериментальних точок і точок лінії регресії при відповідних значеннях змінної x. Тому метод, основою якого є критерій якості у вигляді (5.33), носить назву *метода найменших квадратів.* Рівняння (5.33) є функцією параметрів моделі, які підлягають визначенню.

5.4 Побудова лінійної регресійної моделі

5.4.1 Знаходження коефіцієнтів лінійного рівняння регресії за допомогою МНК

Як відомо, умовний екстремум функції декількох змінних визначається шляхом зрівнювання до нуля частинних похідних цієї функції по незалежних змінних. Нехай, наприклад,

$$\bar{y}(x) = f(a, b, x), \tag{5.34}$$

тобто функція регресії має два невідомих параметра a і b. Умови екстремуму визначаються рівняннями:

$$\frac{\partial \Delta^2}{\partial a} = 0; \qquad \frac{\partial \Delta^2}{\partial b} = 0. \tag{5.35}$$

Розраховані за допомогою умов (5.35) параметри a і b надають дійсно мінімум критерію якості Δ^2. Як буде показано далі, операції (5.35) приводять до системи алгебраїчних рівнянь, розв'язок якої дає шукані значення параметрів моделі. Ці *алгебраїчні рівняння називають нормальними*, а відповідну систему *системою нормальних рівнянь*.

Параметри вибраної регресійної моделі знаходять на основі вибірок випадкових величин X та Y. Тому вони є оцінками параметрів генерального рівняння регресії. Якщо вони отримані за методом найменших квадратів, то їх називають оцінками метода найменших квадратів.

Як відомо, лінійне рівняння регресії має вигляд:

$$\bar{y}(x) = ax + b, \tag{5.36}$$

де a і b – коефіцієнти регресії, які треба визначити на основі вибірок випадкових величин X та Y.

Для їх знаходження застосуємо метод найменших квадратів. Відповідний критерій якості має таку форму:

$$\Delta^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 = min. \tag{5.37}$$

Щоб отримати систему нормальних рівнянь, прирівняємо до нуля частинні похідні по a і b від критерію Δ^2:

$$\frac{\partial \Delta^2}{\partial a} = -2\sum_{i=1}^{n} (y_i - ax_i - b)x_i = 0; \tag{5.38}$$

$$\frac{\partial \Delta^2}{\partial b} = -2\sum_{i=1}^{n}(y_i - ax_i - b) = 0. \tag{5.39}$$

Після простих перетворень ми приходимо до системи алгебраїчних рівнянь:

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \tag{5.40}$$

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i, \tag{5.41}$$

які дозволяють отримати спочатку значення коефіцієнта a:

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}. \tag{5.42}$$

Індекси сумування в рівнянні (5.42) тимчасово опущені.

Розділимо чисельник і знаменник рівняння (5.42) на n^2. Будемо мати :

$$a = \frac{\frac{1}{n}\sum x_i y_i - \frac{1}{n}\sum x_i \frac{1}{n}\sum y_i}{\frac{1}{n}\sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2}$$

або

$$a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - (\overline{x})^2}. \tag{5.43}$$

Виконуючи ряд математичних перетворень отримаємо такі рівняння:

$$\overline{xy} - \overline{x}\cdot\overline{y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = K_{xy}, \tag{5.44}$$

$$\overline{x^2} - (\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = S_x^2. \tag{5.45}$$

K_{xy} *– носить назву коваріації випадкових величин* X та Y.

Відомо, що

$$K_{xy} = r_{xy} S_x S_y, \tag{5.46}$$

де r_{xy} *– статистичний параметр, який має назву коефіцієнта кореляції.*

Враховуючи рівності (5.44)–(5.46), отримаємо формулу для обчислення кутового коефіцієнта лінійного рівняння регресії:

$$a = r_{xy}\frac{S_y}{S_x} . \qquad (5.47)$$

Аналогічно, для коефіцієнта регресії b маємо:

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} . \qquad (5.48)$$

Формулу (5.48) можна переписати таким чином:

$$b = \frac{n\sum x_i^2 \frac{\sum y_i}{n} - \frac{\sum x_i}{n} n\sum x_i y_i + \frac{1}{n}(\sum x_i)^2 \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} - \frac{\frac{1}{n}(\sum x_i)^2 \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} =$$

$$= \overline{y}\frac{n\sum x_i^2 - (\sum x_i)^2}{n\sum x_i^2 - (\sum x_i)^2} - \overline{x}\frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

або, після врахування формули (5.42) та скорочень, маємо:

$$b = \overline{y} - a\overline{x} . \qquad (5.49)$$

Оскільки коефіцієнти a і b лінійного рівняння регресії отримані за методом найменших квадратів, їх називають оцінками метода найменших квадратів.

5.4.2 Перевірка гіпотези про статистичну значущість коефіцієнтів лінійного рівняння регресії

Після знаходження коефіцієнтів лінійної регресії на основі випадкових вибірок необхідно перевірити гіпотези про статистичну значущість a і b.

Гіпотеза основна та альтернативна формулюються так:

H_0 : На заданому рівні значущості α $A = 0$; $B = 0$;

відповідно $\hat{A} = a = 0$; $\hat{B} = b = 0$.

H_1 : На заданому рівні значущості α $A \neq 0$; $B \neq 0$;

відповідно $\hat{A} = a \neq 0$; $\hat{B} = b \neq 0$.

Гіпотези про статистичну значущість коефіцієнтів лінійної регресійної моделі перевіряють на рівні значущості α за допомогою критерію Стьюдента, аналогічно тому, як це проводилося для коефіцієнта кореляції.

При цьому:

$$t = \frac{|a|}{\sigma_a}, \tag{5.50}$$

$$t = \frac{|b|}{\sigma_b}. \tag{5.51}$$

Стандартні відхили коефіцієнтів регресії σ_a та σ_b дорівнюють:

$$\sigma_a = \frac{S_y}{S_x \sqrt{n}}, \tag{5.52}$$

$$\sigma_b = \frac{S_y}{\sqrt{n}} \left[1 + \frac{1}{C_{V_x}^2} \right]^{1/2}, \tag{5.53}$$

де

$$C_{V_x} = \frac{S_x}{|\bar{x}|} \text{ – коефіцієнт варіації.} \tag{5.54}$$

Висновок про статистичну незначущість того чи іншого коефіцієнта регресії роблять у тому випадку, коли виявляється, що $t < t_{кр}(\alpha, \nu)$.

У протилежному випадку, тобто, якщо $t > t_{кр}(\alpha, \nu)$, відповідний коефіцієнт регресії на рівні значущості α є статистично значущим. А це дає підстави виразити кореляційний зв'язок між двома випадковими величинами X та Y лінійним рівнянням регресії $\bar{y}(x) = ax + b$.

Запитання для самоперевірки

1. Які види зв'язків можуть спостерігатися між двома випадковими величинами?
2. Яка залежність між випадковими величинами називається функціональною? стохастичною?
3. Дайте визначення кореляційної залежності між двома випадковими величинами.

4. В якому випадку кореляційна залежність між випадковими величинами буде функціональною?
5. Який розподіл називається умовним? Що є умовним математичним сподіванням? умовною дисперсією?
6. Що є якісною характеристикою тісноти та форми кореляційного зв’язку між двома випадковими величинами ?
7. Які форми кореляційного зв’язку між двома випадковими величинами Вам відомі?
8. Який вигляд має рівняння лінійної регресії та який сенс його коефіцієнтів?
9. Які види нелінійних рівнянь регресії між випадковими величинами Вам відомі?
10. Що є кількісною мірою лінійного кореляційного зв’язку між двома випадковими величинами ?
11. Які статистичні оцінки і яких моментів розподілу випадкових величин використовуються при розрахунках коефіцієнта кореляції?
12. У яких межах змінюється коефіцієнт кореляції?
13. Зобразити графік прямого, тісного, лінійного кореляційного зв’язку.
14. В якому випадку лінійний кореляційний зв'язок буде тіснішим: за умови $r_{xy} = -0{,}82$ чи $r_{xy} = 0{,}82$?
15. Який буде зв'язок між двома випадковими величинами за умови $r_{xy} = 0$?
16. Який буде зв'язок між двома випадковими величинами за умови $\left|r_{xy}\right| = 1$?
17. Що називається «статистичною гіпотезою» та в чому полягає основний принцип її перевірки?
18. За допомогою якого критерію перевіряється гіпотеза про статистичну значущість коефіцієнта кореляції?
19. Як розрахувати фактичне значення критерію Стьюдента для перевірки гіпотези про значущість коефіцієнта кореляції за умови великих (малих) об’ємів вибірок? У якому випадку коефіцієнт кореляції буде статистично значущим?
20. В якому випадку слід будувати лінійне рівняння регресії, що відбиває кореляційну залежність між двома випадковими величинами?
21. Будувати чи ні лінійне рівняння регресії між випадковими величинами X та Y, якщо коефіцієнт кореляції r_{xy} дорівнює 0,48?
22. Який сенс рівня значущості та як визначається ця величина?
23. Дати визначення довірчої ймовірності та який вона має зв’язок з помилкою першого роду?
24. Як визначаються критичні області та області прийняття статистичних гіпотез?

25. Який метод використовується при розрахунках коефіцієнтів рівнянь регресії? Що є «критерієм якості» регресійної моделі?
26. Записати рівняння, що характеризує метод найменших квадратів та пояснити його сенс.
27. Як розрахувати кутовий коефіцієнт лінійного рівняння регресії та його вільний член?
28. За допомогою якого критерію перевіряється статистична гіпотеза про значущість коефіцієнтів лінійного рівняння регресії?
29. Від яких величин залежить критичне значення критерію Стьюдента при перевірках статистичних гіпотез про значущість коефіцієнтів лінійного рівняння регресії?
30. За яких умов статистичні оцінки коефіцієнтів лінійного рівняння регресії будуть значущими та з якою ймовірністю можна це стверджувати?
31. Які види нелінійних рівнянь регресії зустрічаються при дослідженнях статистичної залежності між двома випадковими величинами? Навести приклади.

6. ІНТЕРВАЛЬНІ ОЦІНКИ ПАРАМЕТРІВ

6.1 Поняття про довірчий інтервал

Статистичні оцінки параметрів генеральної сукупності, що розглядалися у попередніх розділах, називають *точковими оцінками*. Така їх назва обумовлюється тим, що в результаті їх розрахунків на основі вибірки, ми отримуємо відповідне число. Як відомо, кожному дійсному числу відповідає точка на числовій осі.

Точкові оцінки того чи іншого параметра генеральної сукупності є *випадковими величинами*. Їх випадковість обумовлюється тим, що вони отримані на основі випадково вилученої з генеральної сукупності вибірки. Тому неможливо зробити висновок про те, у якій мірі оцінка параметра, що отримана на основі даної вибірки, відрізняється від параметра генеральної сукупності, нам, взагалі кажучи, невідомого. Інша річ, коли ми з визначеною ймовірністю можемо отримати числовий інтервал, в якому розташовується параметр генеральної сукупності. Це дає можливість прийти до висновку про якість статистичної оцінки параметра. Ясно, що оцінка тим краща, чим більш вузьким є цей інтервал, і навпаки. Такий інтервал називають *довірчим інтервалом*, а процес побудови його називають *інтервальним оцінюванням параметра генеральної сукупності.*

Довірчим інтервалом $[\theta_1;\theta_2]$ для параметра θ називається такий інтервал, відносно котрого можна з наперед вибраною ймовірністю, близькою до одиниці $(p=1-\alpha)$, стверджувати, що він утримує невідоме значення параметра θ. Це визначення можна записати так:

$$P[\theta_1<\theta<\theta_2]=1-\alpha\,. \qquad (6.1)$$

Оскільки межі довірчого інтервалу θ_1 і θ_2 залежать від елементів вибірки, то їх значення можуть змінюватися від вибірки до вибірки.

Ймовірність $p=1-\alpha$ називають *довірчою ймовірністю,* а α – *рівнем значущості*, який задається дослідником.

6.2 Довірчий інтервал для математичного сподівання, дисперсії та середнього квадратичного відхилу випадкової величини

Точковою оцінкою "$\wedge$" *математичного сподівання* m_x *випадкової величини* X є, як відомо, *середнє значення цієї величини* $(\hat{m}_X=\overline{x})$. Довірчий інтервал для математичного сподівання визначається такою нерівністю:

$$\overline{x} - t_{кр}(\alpha,\nu)\,\sigma_{\overline{x}} < m_x < \overline{x} + t_{кр}(\alpha,\nu)\sigma_{\overline{x}} \;, \tag{6.2}$$

де

$\sigma_{\overline{x}}$ – середня квадратична похибка середнього значення випадкової величини *X*.

Вона визначається формулою:

$$\sigma_{\overline{x}} = \frac{S_x}{\sqrt{n}} \;, \tag{6.3}$$

в якій S_x – середній квадратичний відхил випадкової величини X .

Якщо підставити рівність (6.3) у нерівність (6.2), то будемо мати вираз для інтервального оцінювання математичного сподівання випадкової величини *X:*

$$\overline{x} - t_{кр}(\alpha,\nu)\frac{S_x}{\sqrt{n}} < m_x < \overline{x} + t_{кр}(\alpha,\nu)\frac{S_x}{\sqrt{n}} \;. \tag{6.4}$$

У нерівностях (6.2) та (6.4) $t_{кр}(\alpha,\nu)$ – критичне значення критерію Стьюдента за умов рівня значущості α і числа степенів вільності $\nu = n$, де n – об’єм вибірки (Додаток А).

Оскільки статистична оцінка дисперсії випадкової величини *X* (S_x^2) має χ^2- розподіл, то у загальному випадку довірчий інтервал для дисперсії генеральної сукупності σ_x^2 будують саме за допомогою цього розподілу:

$$\frac{S_x^2 n}{\chi_2^2} < \sigma_x^2 < \frac{S_x^2 n}{\chi_1^2} \;, \tag{6.5}$$

де

n – об’єм вибірки;

χ_1^2 і χ_2^2 – відповідні критичні точки χ^2- розподілу знаходять по Таблицях (Додаток В).

Точка χ_2^2 визначається за умов рівня значущості $\frac{\alpha}{2}$ та числа степенів вільності $\nu = n-1$, а точка χ_1^2 – за умов того ж числа степенів вільності та рівня значущості $\left(1-\frac{\alpha}{2}\right)$.

Для середнього квадратичного відхилу випадкової величини *X* довірчий інтервал має вигляд:

$$\sqrt{\frac{n\,S_x^2}{\chi_2^2}} < \sigma_x < \sqrt{\frac{n\,S_x^2}{\chi_1^2}}\ . \qquad (6.6)$$

За умови великих значень числа степенів вільності розподіл $f\left(\chi^2\right)$ наближається до нормального розподілу. Отже, за такої умови оцінка S_x має розподіл, близький до нормального, з параметрами $M\left[S_x\right] = \sigma_x$ і $\sigma_{S_x}^2 = \dfrac{S_x^2}{2(n-1)}$.

Звідки

$$\sigma_{S_x} = \frac{S_x}{\sqrt{2\left(n-1\right)}} \qquad (6.7)$$

і довірчий інтервал для середнього квадратичного відхилу випадкової величини X *за умови великих об'ємів вибірки* має вигляд:

$$S_x - t_{кр}\left(\alpha,\nu\right)\sigma_{S_x} < \sigma_x < S_x + t_{кр}\left(\alpha,\nu\right)\sigma_{S_x}\,. \qquad (6.8)$$

Для побудови в цьому разі довірчого інтервалу для дисперсії генеральної сукупності σ_x^2 треба підвести до квадрату всі частини нерівності (6.8):

$$\left[S_x - t_{кр}\left(\alpha,\nu\right)\sigma_{S_x}\right]^2 < \sigma_x^2 < \left[S_x + t_{кр}\left(\alpha,\nu\right)\sigma_{S_x}\right]^2. \qquad (6.9)$$

6.3 Довірчий інтервал для коефіцієнта кореляції

Статистична оцінка коефіцієнта кореляції, що розрахована за формулою (5.23), є точковою оцінкою генерального коефіцієнта кореляції ρ_{xy}. Для обгрунтування вірогідності такої оцінки потрібно побудувати довірчий інтервал.

Коефіцієнт кореляції r_{xy}, особливо коли *об'єми* вибірок *невеликі*, не підпорядковується нормальному закону. Але логарифмічне z - перетворення

$$\hat{z} = \frac{1}{2}\ln\frac{1+r_{xy}}{1-r_{xy}} \qquad (6.10)$$

має розподіл, близький до нормального, з параметрами:

$$z = \frac{1}{2}\ln\frac{1+\rho_{xy}}{1-\rho_{xy}} + \frac{\rho_{xy}}{2(n-1)} \tag{6.11}$$

та

$$\sigma_z^2 = \frac{1}{n-3} \ . \tag{6.12}$$

Цей факт дає можливість таким же чином побудувати довірчий інтервал для параметра z, як і для математичного сподівання нормально розподіленої випадкової величини (див. підрозділ 6.2), а саме:

$$\hat{z} - t_{кр}(\alpha,\nu)\sigma_z < z < \hat{z} + t_{кр}(\alpha,\nu)\sigma_z \tag{6.13}$$

або

$$z_1 < z < z_2 \ , \tag{6.14}$$

де

$$z_1 = \hat{z} - t_{кр}(\alpha,\nu)\sigma_z \ , \tag{6.15}$$

$$z_2 = \hat{z} + t_{кр}(\alpha,\nu)\sigma_z \ , \tag{6.16}$$

$$\sigma_z = \frac{1}{\sqrt{n-3}} \, . \tag{6.17}$$

Після обчислення z_1 і z_2 на заданому рівні значущості α та числі степенів вільності $\nu = n$ треба для меж z_1 і z_2 отримати відповідні довірчі межі для коефіцієнта кореляції, тобто r_{xy_1} та r_{xy_2} (Додаток К). Для цього необхідно скористатися оберненим перетворенням. Воно має вигляд:

$$r_{xy_i} = \frac{e^{2z_i} - 1}{e^{2z_i} + 1} \qquad (i = 1,2) \ . \tag{6.18}$$

Отже, довірчий інтервал для коефіцієнта кореляції визначається нерівністю:

$$r_{xy_1} < \rho_{xy} < r_{xy_2} \ . \tag{6.19}$$

Якщо об'єми вибірок, на основі яких оцінюється коефіцієнт кореляції r_{xy}, *є великими*, то оцінка коефіцієнта кореляції має розподіл, близький до нормального з параметрами ρ_{xy} і σ_r^2. Довірчий інтервал за таких умов можна побудувати за допомогою нерівності:

$$r_{xy} - t_{кр}(\alpha,\nu)\sigma_r < \rho_{xy} < r_{xy} + t_{кр}(\alpha,\nu)\sigma_r \, . \tag{6.20}$$

Стандартний відхил коефіцієнта кореляції σ_r розраховується за формулою (6.21):

$$\sigma_r = \frac{1-r_{xy}^2}{\sqrt{n-1}} \quad . \tag{6.21}$$

6.4 Довірчий інтервал для коефіцієнтів лінійного рівняння регресії

Коефіцієнти a і b лінійного рівняння регресії

$$\bar{y}(x) = ax + b \tag{6.22}$$

є точковими оцінками коефіцієнтів A і B генерального рівняння

$$m_{y/x} = Ax + B\,. \tag{6.23}$$

Для останніх також є сенс отримати інтервальні оцінки.

Довірчий інтервал для коефіцієнта A будують за такою формулою:

$$a - t_{кр}(\alpha,\nu)\frac{S_y}{S_x\sqrt{n}} < A < a + t_{кр}(\alpha,\nu)\frac{S_y}{S_x\sqrt{n}}\,. \tag{6.24}$$

У цій нерівності множник

$$\sigma_a = \frac{S_y}{S_x\sqrt{n}} \tag{6.25}$$

є *стандартним відхилом кутового коефіцієнта регресії*, а $t_{кр}(\alpha,\nu)$ – критичне значення параметра Стьюдента за умов рівня значущості α та числа степенів вільності $\nu = n$ (Додаток А).

Нерівність (6.26) дає можливість отримати довірчий інтервал для коефіцієнта B:

$$b - t_{кр}(\alpha,\nu)\frac{S_y}{\sqrt{n}}\left[1+\frac{1}{C_{\mathrm{v}_x}^2}\right]^{1/2} < B < b + t_{кр}(\alpha,\nu)\frac{S_y}{\sqrt{n}}\left[1+\frac{1}{C_{\mathrm{v}_x}^2}\right]^{1/2}. \tag{6.26}$$

У нерівності (6.26) множник

$$\sigma_в = \frac{S_y}{\sqrt{n}}\left[1+\frac{1}{C_{\mathrm{v}_x}^2}\right]^{1/2} \tag{6.27}$$

є стандартним відхилом вільного члена рівняння регресії.

Параметр

$$C_{\mathrm{v}_x} = \frac{S_x}{|\overline{x}|} \tag{6.28}$$

є *коефіцієнтом варіації* випадкової величини X, який характеризує мінливість цієї величини.

6.5 Довірчий коридор для лінійного рівняння регресії

Інтервальні оцінки коефіцієнтів лінійної регресії характеризують точність їх оцінювання. У деяких випадках ставиться задача окреслити ту смугу, в межах якої з визначеною довірчою ймовірністю p розташовується лінія регресії, що відповідає цьому рівнянню. Якщо ця смуга достатньо вузька, то загальна оцінка рівняння регресії задовольняє дослідника. У протилежному випадку, до результатів розрахунків, що робляться по такому рівнянню регресії, відносяться критично.

Довірчий коридор для лінійного рівняння регресії будують за допомогою такої нерівності:

$$\begin{gathered} ax + b - t_{кр}(\alpha,\nu)\frac{S_y}{\sqrt{n}}\left[1+\frac{(x-\overline{x})^2}{S_y^2}\right]^{1/2} < Ax + B < \\ < ax + b + t_{кр}(\alpha,\nu)\frac{S_y}{\sqrt{n}}\left[1+\frac{(x-\overline{x})^2}{S_y^2}\right]^{1/2} \end{gathered}. \tag{6.29}$$

Надаючи в нерівності (6.29) змінній x послідовно її значення з інтервалу $[x_{\min}; x_{\max}]$, отримаємо верхню (права частина нерівності) і нижню (ліва частина нерівності) межі довірчого коридору. Таким чином, будемо мати смугу, у межах якої з довірчою ймовірністю p розташовується генеральне рівняння регресії.

Формули, що наводяться вище, можна використати і для отримання уявлень про якість експоненціального та гіперболічного рівнянь регресії. Але для цього треба за допомогою логарифмічного перетворення звести ці рівняння регресії до лінійних.

Запитання для самоперевірки

1. У чому полягає фізичний сенс математичного сподівання випадкової величини та з яким моментом розподілу воно має зв'язок?
2. Яка статистична оцінка параметра генеральної сукупності називається «точковою» та яку кількість точкових статистичних оцінок можна отримати з однієї генеральної сукупності?
3. Що називається «довірчим інтервалом» параметра генеральної сукупності?
4. Який сенс рівня значущості та довірчої ймовірності?
5. Як знайти інтервальні оцінку математичного сподівання випадкової величини?
6. Як знайти довірчий інтервал для дисперсії та середнього квадратичного відхилу випадкової величини за умови невеликих $(n < 50)$ та великих $(n > 50)$ об'ємів вибірок?
7. Який параметр визначає кількісну міру тісноти та форми лінійного (нелінійного) кореляційного зв'язку?
8. Які статистичні оцінки та яких моментів розподілу необхідно знайти, щоб розрахувати коефіцієнт кореляції?
9. В яких межах змінюється коефіцієнт кореляції? За умови яких значень коефіцієнта кореляції лінійний кореляційний зв'язок буде прямим? оберненим?
10. Як знайти інтервальну оцінку коефіцієнта кореляції за умови великих (малих) об'ємів вибірок?
11. В якому випадку використовується логарифмічне z–перетворення Фішера при побудові довірчого інтервалу для коефіцієнта кореляції?
12. Як знайти критичні значення критерію Стьюдента для інтервального оцінювання коефіцієнта кореляції?
13. Який вигляд має рівняння регресії, що характеризує лінійний кореляційний зв'язок між двома випадковими величинами?
14. Який параметр називається «коефіцієнтом варіації»?
15. Який сенс діаграми розсіювання точок при дослідженні кореляційної залежності між двома випадковими величинами?
16. Як побудувати довірчий коридор для лінійного рівняння регресії?

Додаток А

Значення критерію Стьюдента для рівня значущості α і числа степенів вільності ν

ν	Двостороння критична область, α							
	0,2	**0,1**	**0,05**	**0,02**	**0,01**	**0,005**	**0,002**	**0,001**
6	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
12	1,36	1,78	2,20	2,68	3,05	3,43	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
20	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	1,32	1,71	2,06	2,49	2,90	3,09	3,47	3,75
25	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,44	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
29	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	2,85	3,16	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29

Додаток Б

Значення критерію Фішера F
для рівня значущості 0,05
ν_1 – число степенів вільності для більшої дисперсії
ν_2 – число степенів вільності для меншої дисперсії

ν_2	ν_1							
	15	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
10	2,845	2,774	2,737	2,700	2,661	2,621	2,580	2,538
11	2,179	2,646	2,609	2,571	2,531	2,490	2,448	2,405
12	2,617	2,544	2,506	2,466	2,426	2,384	2,341	2,296
13	2,533	2,459	2,420	2,380	2,339	2,297	2,252	2,206
14	2,463	2,388	2,349	2,308	2,266	2,223	2,178	2,131
15	2,404	2,328	2,288	2,247	2,204	2,160	2,114	2,066
16	2,352	2,276	2,235	2,194	2,151	2,106	2,059	2,010
17	2,308	2,230	2,190	2,148	2,104	2,058	2,010	1,960
18	2,269	2,191	2,150	2,107	2,063	2,017	1,968	1,917
19	2,234	2,156	2,114	2,071	2,026	1,980	1,930	1,878
20	2,203	2,124	2,083	2,039	1,994	1,946	1,896	1,843
21	2,176	2,096	2,054	2,010	1,965	1,917	1,866	1,812
22	2,151	2,071	2,028	1,984	1,938	1,890	1,838	1,783
23	2,128	2,048	2,005	1,961	1,914	1,865	1,813	1,757
24	2,108	2,027	1,984	1,939	1,892	1,842	1,790	1,733
25	2,089	2,008	1,964	1,919	1,872	1,822	1,768	1,711
26	2,072	1,990	1,946	1,901	1,853	1,803	1,749	1,691
27	2,056	1,974	1,930	1,884	1,836	1,785	1,731	1,672
28	2,041	1,959	1,915	1,869	1,820	1,769	1,714	1,654
29	2,028	1,945	1,901	1,854	1,806	1,754	1,698	1,638
30	2,015	1,932	1,987	1,841	1,792	1,740	1,684	1,622
40	1,925	1,839	1,793	1,744	1,693	1,637	1,577	1,509
60	1,836	1,748	1,700	1,649	1,594	1,534	1,467	1,389
120	1,751	1,659	1,608	1,554	1,495	1,429	1,352	1,254
∞	1,666	1,571	1,517	1,459	1,394	1,318	1,221	1,000

Додаток В

Значення $\chi^2(\alpha,\nu)$ для різних чисел степенів вільності ν та рівня значущості α

ν \ α	0,995	0,990	0,975	0,95	0,90	0,80	0,70	0,30
1	0,0393	0,0157	0,0982	0,0393	0,0158	0,0642	0,148	1,07
2	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	2,41
3	0,0717	0,115	0,216	0,352	0,584	1,00	1,42	3,67
4	0,207	0,297	0,484	0,711	1,06	1,65	2,19	4,88
5	0,412	0,554	0,831	1,15	1,61	2,34	3,00	6,06
6	0,676	0,872	1,24	1,64	2,20	3,07	3,83	7,23
7	0,989	1,24	1,69	2,17	2,83	3,82	4,67	8,38
8	1,34	1,65	2,18	2,73	3,49	4,59	5,53	9,52
9	1,73	2,09	2,70	3,33	4,17	5,38	6,39	10,7
10	2,16	2,56	3,25	3,94	4,87	6,18	7,27	11,8
11	2,60	3,05	3,82	4,57	5,58	6,99	8,15	12,9
12	3,07	3,57	4,40	5,23	6,30	7,81	9,03	14,0
13	3,57	4,11	5,01	5,89	7,04	8,63	9,93	15,1
14	4,07	4,66	5,63	6,57	7,79	9,47	10,8	16,2
15	4,60	5,23	6,26	7,26	8,55	10,3	11,7	17,3
16	5,14	5,81	6,91	7,96	9,31	11,2	12,6	18,4
17	5,70	6,41	7,56	8,67	10,1	12,0	13,5	19,5
18	6,26	7,01	8,23	9,39	10,9	12,9	14,4	20,6
19	6,84	7,63	8,91	10,1	11,7	13,7	15,4	21,7
20	7,43	8,26	9,59	10,9	12,4	14,6	16,3	22,8
21	8,03	8,90	10,3	11,6	13,2	15,4	17,2	23,9
22	8,64	9,54	11,0	12,3	14,0	16,3	18,1	24,9
23	9,26	10,2	11,7	13,1	14,8	17,2	19,0	26,0
24	9,89	10,9	12,4	13,8	15,7	18,1	19,9	27,1
25	10,5	10,5	13,1	14,6	16,5	18,9	20,9	28,2
26	11,2	12,2	13,8	15,4	17,3	19,8	21,8	29,2
27	11,8	12,9	14,6	16,2	18,1	20,7	22,7	30,3

Продовження додатку В

α \ ν	0,995	0,990	0,975	0,95	0,90	0,80	0,70	0,30
28	12,5	13,6	15,3	16,9	18,9	21,6	23,6	31,4
29	13,1	14,3	16,0	17,7	19,8	22,5	24,6	32,5
30	13,8	15,0	16,8	18,5	20,6	23,4	25,5	33,5
35	17,2	18,5	20,6	22,5	24,8	27,8	30,2	38,9
40	20,7	22,2	24,4	26,5	29,1	32,3	34,9	44,2
45	24,3	25,9	28,4	30,6	33,4	36,9	39,6	49,5
50	28,0	29,7	32,4	34,8	37,7	41,1	44,3	54,7
75	47,2	49,5	52,9	56,1	59,8	64,5	68,1	80,9
100	67,3	70,1	74,2	77,9	82,4	87,9	92,1	106,9

Продовження додатку В

α \ ν	0,20	0,10	0,05	0,025	0,010	0,005	0,001
1	1,64	2,71	3,84	5,02	6,63	7,88	10,8
2	3,22	4,61	5,99	7,38	9,21	10,6	13,8
3	4,64	6,25	7,81	9,35	11,3	12,8	16,3
4	5,99	7,78	9,49	11,1	13,3	14,9	18,5
5	7,29	9,24	11,1	12,8	15,1	16,7	20,5
6	8,56	10,6	12,6	14,4	16,8	18,5	22,5
7	9,80	12,0	14,1	16,0	18,5	20,3	24,3
8	11,0	13,4	15,5	17,5	20,1	22,0	26,1
9	12,2	14,7	16,9	19,0	21,7	23,6	27,9
10	13,4	16,0	18,3	20,5	23,2	25,2	29,6
11	14,6	17,3	19,7	21,9	24,7	26,8	31,3
12	15,8	18,5	21,0	23,3	26,2	28,3	32,9
13	17,0	19,8	22,4	24,7	27,7	29,8	34,5
14	18,2	21,1	23,7	26,1	29,1	31,3	36,1
15	19,3	22,3	25,0	27,6	30,9	32,8	37,7

Продовження додатку В

α / ν	0,20	0,10	0,05	0,025	0,010	0,005	0,001
16	20,5	23,5	26,3	28,8	32,0	34,3	39,3
17	21,6	24,8	27,6	30,2	33,4	35,7	40,8
18	22,8	26,0	28,9	31,5	34,8	37,2	42,3
19	23,9	27,0	30,1	32,9	36,2	38,6	43,8
20	25,0	28,4	31,4	34,2	37,6	40,0	45,3
21	26,9	29,6	32,7	35,5	38,9	41,4	46,8
22	27,3	30,8	33,9	36,0	40,3	42,8	48,3
23	28,4	32,0	35,2	38,1	41,6	44,2	49,7
24	29,6	33,0	36,4	39,4	43,0	45,6	51,2
25	30,7	34,4	37,7	40,6	44,3	46,9	52,6
26	31,8	35,6	38,9	41,0	45,6	48,3	54,1
27	32,9	36,7	40,1	43,2	47,0	49,6	55,5
28	34,0	37,9	41,3	44,5	48,3	51,0	56,9
29	35,1	39,1	42,6	45,7	49,6	52,3	58,3
30	36,3	40,3	43,8	47,0	50,9	53,7	59,7
35	41,8	46,1	49,8	53,2	57,3	60,3	66,6
40	47,3	51,8	55,8	59,3	63,7	66,8	73,4
45	52,7	57,5	61,7	65,4	70,0	73,2	80,1
50	58,2	63,2	67,5	71,4	76,2	79,5	86,7
75	85,1	91,1	96,2	100,8	106,4	110,3	118,6
100	111,7	118,5	124,3	129,6	135,6	140,2	149,4

Додаток Г

Значення функції $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

t	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3700	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315

Продовження додатку Г

t	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034

Додаток Д

Значення інтеграла ймовірності $\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0,00	0,00000	0,30	0,23582	0,60	0,45149	0,90	0,63188
01	00798	31	24344	61	45814	91	63718
02	01596	32	25103	62	46474	92	64243
03	02393	33	25860	63	47131	93	64763
04	03191	34	26614	64	47783	94	65278
0,05	0,03988	0,35	0,27366	0,65	0,48431	0,95	0,65789
06	04784	36	28115	66	49075	96	66294
07	05581	37	28862	67	49714	97	66795
08	06376	38	29605	68	50350	98	67291
09	07171	39	30346	69	50981	99	67783
0,10	0,07966	0,40	0,31084	0,70	0,51607	1,00	0,68269
11	08759	41	31819	71	52230	01	68750
12	09552	42	32552	72	52848	02	69227
13	10348	43	33280	73	53461	03	69699
14	11134	44	34006	74	54070	04	70166
15	11924	45	34729	75	54675	1,05	70628
16	12712	46	35448	76	55275	06	71086
17	13499	47	36164	77	55870	07	71538
18	14285	48	36877	78	56461	08	71986
19	15069	49	37587	79	57047	09	72429
0,20	0,15852	0,50	0,38292	0,80	0,57629	1,10	0,72867
21	16633	51	38995	81	58206	11	73300
22	17413	52	39694	82	58778	12	73729
23	18191	53	40389	83	59346	13	74152
24	18967	54	41080	84	59909	14	74571
25	19741	55	41768	85	60468	15	74986
26	20514	56	42452	86	61021	16	75395
27	21284	57	43132	87	61570	17	75800
28	22052	58	43809	88	62114	18	76200
29	22818	59	44481	89	62653	19	76595

Продовження додатку Д

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
1,20	0,76986	1,55	0,87886	1,90	0,94257	2,25	0,97555
21	77372	56	88124	91	94387	26	97618
22	77754	57	88358	92	94514	27	97679
23	78130	58	88589	93	94639	28	97739
24	78502	59	88817	94	94762	29	97798
25	78870	1,60	0,89040	95	94882	2,30	0,97855
26	79233	61	89260	96	95000	31	97911
27	79592	62	89477	97	95116	32	97966
28	79945	63	89690	98	95230	33	98019
29	80295	64	89899	99	95341	34	98072
1,30	0,80640	65	90106	2,00	0,95450	35	98123
31	80980	66	90309	01	95557	36	98172
32	81316	67	90508	02	95662	37	98221
33	81648	68	90704	03	95764	38	98269
34	81975	69	90897	04	95865	39	98315
35	82298	1,70	0,91087	05	95964	2,40	0,98360
36	82617	71	91273	06	96060	41	98405
37	82931	72	91457	07	96155	42	98448
38	83241	73	91637	08	96247	43	98490
39	83547	74	91814	09	96338	44	98531
1,40	0,83849	75	91988	2,10	0,96427	45	98571
41	84146	76	92159	11	96514	46	98611
42	84439	77	92327	12	96599	47	98649
43	84728	78	92492	13	96683	48	98686
44	85013	79	92655	14	96765	49	98723
45	85294	1,80	0,92814	15	96844	2,50	0,98758
46	85571	81	92970	16	96923	51	98793
47	85844	82	93124	17	96999	52	98826
48	86113	83	93275	18	97074	53	98859
49	86378	84	93423	19	97148	54	98891
1,50	0,86639	1,85	0,93569	2,20	0,97219	2,55	0,98923
51	86696	86	93711	21	97289	56	98953
52	87149	87	93852	22	97358	57	98983
53	87398	88	93989	23	97425	58	99012
54	87644	89	94124	24	97491	59	99040

Продовження додатку Д

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
2,60	0,99068	2,95	0,99682	3,30	0,99903	3,65	0,99974
61	99095	96	99692	31	99907	66	99975
62	99121	97	99702	32	99910	67	99976
63	99146	98	99712	33	99913	68	99977
64	99171	99	99721	34	99916	69	99978
65	99195	3,00	0,99730	35	99919	3,70	0,99978
66	99219	01	99739	36	99922	71	99979
67	0,99241	02	99747	37	99925	72	99980
68	99263	03	99755	38	99928	73	99981
69	99285	04	99763	39	99930	74	99982
2,70	0,99307	05	99771	3,40	0,99933	75	99982
71	99327	06	99779	41	99935	76	99983
72	99347	07	99786	42	99937	77	99984
73	99367	08	99793	43	99940	78	99984
74	99386	09	99800	44	99942	79	99985
75	99404	3,10	0,99806	45	99944	3,80	0,99986
76	99422	11	99813	46	99946	81	99986
77	99439	12	99819	47	99948	82	99987
78	99456	13	99825	48	99950	83	99987
79	99473	14	99831	49	99952	84	99988
2,80	0,99489	15	99837	3,50	99953	85	99988
81	99505	16	99842	51	99955	86	99989
82	99520	17	99848	52	99957	87	99989
83	99535	18	99853	53	99958	88	99990
84	99549	19	99858	54	99960	89	99990
85	99563	3,20	0,99863	55	99961	3,90	0,99990
86	99576	21	99867	56	99963	91	99991
87	99590	22	99872	57	99964	92	99991
88	99602	23	99876	58	99966	93	99992
89	99615	24	99880	59	99967	94	99992
2,90	0,99627	25	99855	3,60	0,99968	95	99992
91	99639	26	99889	61	99969	96	99992
92	99650	27	99892	62	99971	97	99993
93	99661	28	99896	63	99972	98	99993
94	99672	29	99900	64	99973	99	99993

Додаток Е

Значення гамма – функції $\Gamma(x)$

x	$\Gamma(x)$	x	$\Gamma(x)$	x	$\Gamma(x)$	x	$\Gamma(x)$
1,00	1,00000	1,25	0,90640	1,50	0,88623	1,75	0,91906
1,01	0,99433	1,26	0,90440	1,51	0,88659	1,76	0,92137
1,02	0,98884	1,27	0,90250	1,52	0,88704	1,77	0,92376
1,03	0,98355	1,28	0,90072	1,53	0,88757	1,78	0,92623
1,04	0,97844	1,29	0,89904	1,54	0,88818	1,79	0,92877
1,05	0,97350	1,30	0,89747	1,55	0,88887	1,80	0,93138
1,06	0,96874	1,31	0,89600	1,56	0,88964	1,81	0,93408
1,07	0,96415	1,32	0,89464	1,57	0,89049	1,82	0,93685
1,08	0,95973	1,33	0,89338	1,58	0,89142	1,83	0,93969
1,09	0,95546	1,34	0,89222	1,59	0,89243	1,84	0,92261
1,10	0,95135	1,35	0,89115	1,60	0,89352	1,85	0,94561
1,11	0,94740	1,36	0,89018	1,61	0,89468	1,86	0,94869
1,12	0,94359	1,37	0,89931	1,62	0,89592	1,87	0,95184
1,13	0,93993	1,38	0,88854	1,63	0,89724	1,88	0,95507
1,14	0,93642	1,39	0,88785	1,64	0,89864	1,89	0,95838
1,15	0,93304	1,40	0,88726	1,65	0,90012	1,90	0,96177
1,16	0,92980	1,41	0,88676	1,66	0,90167	1,91	0,96523
1,17	0,92670	1,42	0,88636	1,67	0,90330	1,92	0,96877
1,18	0,92373	1,43	0,88604	1,68	0,90500	1,93	0,97240
1,19	0,92089	1,44	0,88581	1,69	0,90678	1,94	0,97610
1,20	0,91817	1,45	0,88566	1,70	0,90864	1,95	0,97988
1,21	0,91558	1,46	0,88560	1,71	0,91057	1,96	0,98374
1,22	0,91311	1,47	0,88563	1,72	0,91258	1,97	0,98768
1,23	0,91075	1,48	0,88575	1,73	0,91467	1,98	0,99171
1,24	0,90852	1,49	0,88595	1,74	0,91683	1,99	0,99581
1,25	0,90640	1,50	0,88623	1,75	0,91906	2,00	1,00000

Значення гамма-функції для $x < 1\,(x \neq 0,\ -1,\ -2, \ldots,)$ та для $x > 2$ розраховуються за допомогою формул:

$$\Gamma(x) = \frac{\Gamma(x+1)}{x}\ ; \qquad\qquad \Gamma(x) = (x-1)\,\Gamma(x-1)\ .$$

Приклади. 1. $\Gamma(0{,}7) = \Gamma(1{,}7) / 0{,}7 = 0{,}90864 / 0{,}7 = 1{,}2981$;

2. $\Gamma(3{,}5) = 2{,}5\ \Gamma(2{,}5) = 2{,}5 \cdot 1{,}5\ \Gamma(1{,}5) =$
$= 2{,}5 \cdot 1{,}5 \cdot 0{,}88623 = 3{,}32336$.

Додаток Ж

Значення функції Пуассона $P(X=i)=\dfrac{\lambda^i e^{-\lambda}}{i\ !}$

i	λ									
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0	0,9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0253
5	-	-	-	0001	0002	0004	0007	0012	0020	0031
6	-	-	-	-	-	-	0001	0002	0003	0005
7	-	-	-	-	-	-	-	-	-	0001

Продовження додатку Ж

i	λ									
	1,1	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**
0	0,3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1805
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0063	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0013	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8	-	-	0001	0001	0001	0002	0003	0005	0006	0009
9	-	-	-	-	-	-	0001	0001	0001	0002

Продовження додатку Ж

i	*λ*									
	2,1	**2,2**	**2,3**	**2,4**	**2,5**	**2,6**	**2,7**	**2,8**	**2,9**	**3,0**
0	0,1225	1108	1003	0903	0821	0743	0672	0608	0550	0498
1	2572	2438	2306	2177	2052	1931	1815	1703	1596	1494
2	2700	2681	2652	2613	2565	2510	2450	2384	2314	2240
3	1890	1964	2083	2090	2136	2176	2205	2225	2234	2240
4	0992	1087	1169	1254	1336	1414	1488	1557	1622	1680
5	0417	0476	0538	0602	0668	0735	0804	0872	0941	1008
6	0146	0175	0206	0241	0278	0319	0362	0407	0455	0504
7	0004	0055	0068	0083	0099	0118	0140	0163	0188	0216
8	0012	0015	0020	0025	0031	0039	0047	0057	0068	0081
9	0003	0004	0005	0007	0009	0011	0014	018	0022	0027
10	0001	0001	0001	0002	0002	0003	0004	0005	0006	0008

Продовження додатку Ж

i	*λ*									
	3,5	**4,0**	**4,5**	**5,0**	**6,0**	**7,0**	**8,0**	**9,0**	**10,0**	**11,0**
0	0,0302	0183	0111	0067	0025	0009	0003	0001	-	-
1	1057	0733	0500	0333	0149	0064	0027	0011	0005	0002
2	1850	1465	1125	0842	0446	0223	0107	0050	0023	0010
3	2158	1954	1687	1404	0892	0521	0286	0150	0076	0037
4	1888	1954	1898	1755	1339	0912	0573	0338	0189	0102
5	1327	1563	1708	1755	1606	1277	0916	1277	0916	0224
6	0771	1042	1281	1462	1606	1490	1221	0911	0631	0411
7	0386	0595	0824	1044	1377	1490	1396	1171	0901	0646
8	0169	0298	0463	0653	1033	1304	1396	1318	1126	0888
9	0066	0132	0232	0363	0638	1014	1241	1318	1251	1058
10	0023	0053	0104	0181	0413	0710	0993	1186	1251	1194
11	0007	0019	0043	0082	0225	0452	0722	0970	1137	1194
12	0002	0006	0016	0034	0113	0264	0481	0728	0948	1094

Додаток К

Значення величини r для значень Z від 0,00 до 2,99

z	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,119	0,129	0,139	0,149	0,159	0,168	0,178	0,188
0,2	0,197	0,207	0,217	0,226	0,236	0,245	0,254	0,264	0,273	0,282
0,3	0,291	0,300	0,310	0,319	0,328	0,336	0,345	0,354	0,363	0,371
0,4	0,380	0,389	0,397	0,405	0,414	0,422	0,430	0,438	0,446	0,454
0,5	0,462	0,470	0,478	0,485	0,493	0,501	0,508	0,515	0,523	0,530
0,6	0,537	0,544	0,551	0,558	0,565	0,572	0,578	0,585	0,592	0,598
0,7	0,604	0,611	0,617	0,623	0,629	0,635	0,641	0,647	0,653	0,658
0,8	0,664	0,670	0,675	0,681	0,686	0,691	0,696	0,701	0,706	0,711
0,9	0,716	0,721	0,726	0,731	0,735	0,740	0,744	0,749	0,753	0,757
1,0	0,762	0,766	0,770	0,774	0,778	0,782	0,786	0,790	0,793	0,697
1,1	0,801	0,804	0,808	0,811	0,814	0,818	0,821	0,824	0,828	0,831
1,2	0,834	0,837	0,840	0,843	0,846	0,848	0,851	0,854	0,857	0,859
1,3	0,862	0,864	0,867	0,869	0,872	0,874	0,876	0,879	0,881	0,883
1,4	0,885	0,888	0,890	0,892	0,894	0,896	0,898	0,900	0,902	0,903
1,5	0,905	0,907	0,909	0,910	0,912	0,914	0,915	0,917	0,919	0,920
1,6	0,922	0,923	0,925	0,926	0,928	0,929	0,930	0,932	0,933	0,934
1,7	0,935	0,937	0,940	0,939	0,940	0,941	0,943	0,944	0,945	0,946

Продовження додатку К

z	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1,8	0,947	0,948	0,949	0,950	0,951	0,952	0,953	0,954	0,955	0,955
1,9	0,956	0,957	0,958	0,959	0,960	0,960	0,961	0,962	0,963	0,963
2,0	0,964	0,965	0,965	0,966	0,967	0,967	0,968	0,969	0,969	0,970
2,1	0,970	0,971	0,972	0,972	0,973	0,937	0,974	0,974	0,975	0,975
2,2	0,976	0,976	0,977	0,977	0,978	0,978	0,979	0,979	0,979	0,980
2,3	0,980	0,981	0,981	0,981	0,982	0,982	0,982	0,983	0,983	0,983
2,4	0,984	0,984	0,984	0,985	0,985	0,985	0,986	0,986	0,986	0,986
2,5	0,987	0,987	0,987	0,987	0,988	0,988	0,988	0,988	0,989	0,989
2,6	0,989	0,989	0,989	0,990	0,990	0,990	0,990	0,990	0,991	0,991
2,7	0,991	0,991	0,991	0,992	0,992	0,992	0,992	0,992	0,992	0,993
2,8	0,993	0,993	0,993	0,993	0,993	0,993	0,994	0,994	0,994	0,994
2,9	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995

Л І Т Е Р А Т У Р А

1. Гончарова Л. Д., Школьний Є. П. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ): Навчальний посібник. / Одеса: Екологія, 2007. 460 с.
2. Груза Г.В., Рейтенбах Р.Г. Статистика и анализ гидрометеорологических данных. / Л.: Гидрометеоиздат, 1982. 216 с.
3. Груза Г.В., Ранькова Э.Я. Вероятностные метеорологические прогнозы. /Л.: Гидрометеоиздат, 1983. 271 с.
4. Исаев А.А. Статистика в метеорологии и климатологии. / М.: Изд. МГУ, 1988. 248 с.
5. Митропольский А.К. Техника статистических вычислений. / М.: Наука, 1961. 479 с.
6. Рождественский А.В., Чеботарев А.И. Статистические методы в гидрологии. /Л.: Гидрометеоиздат, 1974. 424 с.
7. Сикан А.В. Методы статистической обработки гидрометеорологической информации: учебник. / СПб: РГГМУ, 2007. 279 с.
8. Шефтель З.Г. Теорія ймовірностей. / К.: Вища школа, 1994. 192 с.
9. Школьний Є.П., Лоєва І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації: підручник / Одеса: ТЕС, 1999. 600 с.
10. Школьний Є. П., Гончарова Л. Д., Миротворська Н. К. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ): навчальний посібник./ Одеса: ТЕС, 2000. 420 с.

ЗМІСТ