МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

МЕТОДИЧНІ ВКАЗІВКИ
ДО ЛАБОРАТОРНИХ ТА САМОСТІЙНИХ РОБІТ
ІЗ ДИСЦИПЛІНИ
"МАТЕМАТИЧНА СТАТИСТИКА"

ДЛЯ СТУДЕНТІВ МЕХАНІКО-МАТЕМАТИЧНОГО ФАКУЛЬТЕТУ

КИЇВський університет
2014

Рецензенти:
д-р фіз.-мат. наук, проф. Ю. В. Козаченко,
д-р фіз.-мат. наук, проф. Є. О. Лєбєдєв

Рекомендовано до друку вченою радою
механіко-математичного факультету
(протокол № 3 від 8 жовтня 2012 року)

Ухвалено науково-методичною радою
Київського національного університету імені Тараса Шевченка
14 жовтня 2013 року

Методичні вказівки до лабораторних та самостійних робіт із дисципліни "Математична статистика" / упорядники : О. І. Василик, М. В. Карташов, В. П. Кнопова [та ін.]. – К. : Видавничо-поліграфічний центр "Київський університет", 2014. – 84 с.

Видання призначено для студентів третього курсу механіко-математичного факультету спеціальностей "Математика" та "Статистика". Зміст і структура матеріалу відповідають програмі дисципліни та вимогам кредитно-модульної системи організації навчального процесу. Матеріал до кожного заняття містить задачі для аудиторної та самостійної роботи, а також додаткові задачі. Наведено посилання на літературу, де розглянуто теоретичні відомості, необхідні для розв'язування задач.

Зміст

Передмова

Методичні вказівки до лабораторних та самостійних робіт із нормативної дисципліни "Математична статистика" призначені для студентів III курсу механіко-математичного факультету спеціальностей "Математика", "Статистика". Зміст та структура матеріалу відповідають програмі дисципліни та вимогам кредитно-модульної системи організації навчального процесу.

Розділи посібника містять завдання для лабораторних та самостійних робіт, що входять до складу двох модулів дисципліни:

1. Статистичне оцінювання і перевірка гіпотез.

2. Елементи статистики випадкових процесів.

Перший із цих модулів призначений для вивчення основних методів статистичного оцінювання, їх практичного застосування та оволодіння певними методами перевірки статистичних гіпотез. Другий модуль призначений для ознайомлення зі статистикою випадкових процесів.

Завдання робіт із кожної теми розбито на три групи:

A – задачі аудиторних лабораторних робіт,

B – задачі позааудиторних чи самостійних робіт,

C – додаткові задачі.

Задачі групи **A** розв'язуються в аудиторії під керівництвом та за вказівками викладача. Задачі групи **B** є аналогічними за змістом аудиторним задачам. Позааудиторні та самостійні завдання з групи **B** виконуються студентами самостійно на підставі прослуханих лекцій, отриманих у аудиторії навиків для позааудиторних завдань та поточних консультацій із теоретичного матеріалу для самостійних робіт. Завдання з групи **C** є додатковими для можливої заміни завдань з **A, B**. Якість виконання вказаних завдань контролюється викладачами, оцінюється у балах та враховується згідно з кредитно-модульною системою при оцінюванні рівня знань студентів.

У додатку наведено таблиці всіх необхідних стандартних розподілів та їх квантилів. За відсутності даних для проміжних значень параметрів слід застосувати метод лінійної інтерполяції.

Матеріал посібника розбито на теми відповідно до модулів.

Модуль 1.
Статистичне оцінювання і перевірка гіпотез

1.1. Задачі непараметричного оцінювання

1.1.1. Статистичний простір, вибірка, функція вірогідності. Статистики й оцінки. Властивості оцінок

Література: [1, с. 319–327], [2, с. 324–331]

А

1. Протягом року аналізуються повні дані портфеля з n страхових полісів при невідомій імовірності θ настання страхового випадку за припущення незалежності таких випадків для різних полісів. Побудувати відповідний статистичний простір, описати статистичну вибірку. Обчислити функцію вірогідності. Навести приклади оцінок.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з рівномірного розподілу $U(a, b)$, $a < b$. Розглядаються оцінки:

$$\hat{\theta}_1 = \min_{1 \le j \le n} \xi_j,\ \ \hat{\theta}_2 = \max_{1 \le j \le n} \xi_j,\ \hat{\theta}_3 = n^{-1} \sum_{i=1}^{n} \xi_i,\ \hat{\theta}_4 = (\xi_1 + \xi_n)/2.$$

Які з оцінок $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ є а) незміщеними оцінками, б) консистентними оцінками, в) асимптотично нормальними для одного з параметрів a, b або для $(a+b)/2$?

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з показникового розподілу зі щільністю розподілу $f(y, \theta) = \theta^{-1} \exp\left\{-\theta^{-1} y\right\} \mathbb{1}_{\{y>0\}}, \theta > 0$. Довести, що оцінка $\hat{\theta}_1 = n^{-1} \sum\limits_{i=1}^{n} \xi_i$ є незміщеною та консистентною оцінкою параметра θ. Чи є оцінка $\hat{\theta}_2 = (\xi_1 + \xi_n)/2$ а) незміщеною, б) консистентною, в) асимптотично нормальною оцінкою параметра θ?

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з гамма-розподілу $\Gamma(1, \alpha)$. Довести, що статистика $\sum\limits_{j=1}^{k} c_{nkj} \left(\sum\limits_{i=1}^{n} \xi_i \right)^j$ є незміщеною оцінкою параметра α^k при деякій сталій c_{nkj} та знайти значення c_{nkj}.

5. Припустимо, що $\widehat{\theta}$ – незміщена оцінка параметра θ і має дисперсію $\theta^2/10$. Обчислити середньоквадратичну похибку оцінки

$k\widehat{\theta}$, де k – стала, і визначити значення k, для якого ця похибка є мінімальною.

B

1. Проведено n незалежних вимірювань відстані від Землі до Місяця, що ґрунтуються на даних обсерваторних обстежень. Дослідники приймають нормальну (гауссівську) модель для похибок спостережень. Побудувати відповідний статистичний простір, описати статистичну вибірку. Обчислити функцію вірогідності. Навести приклади оцінок.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу зі щільністю розподілу $f(y, \theta) = \exp\{\theta - y\} \mathbb{1}_{\{y>\theta\}}$. Чи є оцінка $\hat{\theta} = -\frac{1}{n} + \min\{\xi_1, \ldots, \xi_n\}$ а) незміщеною, б) консистентною, в) асимптотично нормальною оцінкою параметра θ? Дослідити граничну поведінку цієї оцінки.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу Пуассона з параметром θ. Розглянемо оцінки $\hat{\theta}_1 = n^{-1} \sum\limits_{i=1}^{n} \xi_i$, $\hat{\theta}_2 = n^{-1} \sum\limits_{i=1}^{n} \xi_i^2$. Чи є серед оцінок $\hat{\theta}_1$ та $\hat{\theta}_2$ а) незміщені оцінки, б) консистентні оцінки, в) асимптотично нормальні оцінки параметра θ?

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з рівномірного розподілу на інтервалі $(-\theta, \theta)$. Довести, що оцінка

$$\widehat{\theta} = \max\left(-\min_{1\le k\le n} \xi_k, \max_{1\le k\le n} \xi_k\right) \frac{n+1}{n}$$

є а) незміщеною, б) консистентною оцінкою параметра θ. Дослідити граничну поведінку цієї оцінки.

5. Нехай $\xi_1, \ldots, \xi_n$ – вибірка з результатів послідовності n випробувань Бернуллі з імовірністю успіху θ. Довести, що статистика $U_{ij} = \xi_1 \ldots \xi_i (1 - \xi_{i+1}) \ldots (1 - \xi_{i+j})$ є незміщеною оцінкою для функції $\theta^i (1-\theta)^j$ при $i + j \le n$.

C

1. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з нормального розподілу $N(\mu, \sigma^2)$. Довести, що оцінка $\sum\limits_{k=1}^{n} \left(\xi_k - n^{-1} \sum\limits_{i=1}^{n} \xi_i\right)^2$ є незміщеною оцінкою параметра $(n-1)\sigma^2$.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з нормального розподілу $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Довести, що для деякої сталої c_n оцінка

$$\widehat{\sigma} = c_n \sqrt{\sum_{k=1}^{n} \left(\xi_k - n^{-1} \sum_{i=1}^{n} \xi_i \right)^2}$$

є незміщеною для σ. Знайти цю сталу.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу Парето зі щільністю $f(y, \theta) = ba^{-b} y^{-b-1} \mathbb{1}_{y>a}$, $\theta = (a, b) \in \mathbb{R}_+^2$. Знайти незміщені оцінки для функцій $\tau_1(\theta) = a^{-m}$, при $m \in \mathbb{Z}_+$, $\tau_2(\theta) = b$.

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з гамма-розподілу $\Gamma(\lambda, \alpha)$ із параметром $\theta = (\lambda, \alpha) \in \mathbb{R}_+^2$. Довести, що статистика $n^{-1} \sum\limits_{i=1}^{n} \xi_i$ є а) незміщеною, б) консистентною, в) асимптотично нормальною оцінкою параметричної функції α/λ.

5. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу зі щільністю $f(y, \theta) = a^{-1} \exp\{-(y-b)/a\} \mathbb{1}_{y>b}$, $\theta = (a, b) \in \mathbb{R}_+ \times \mathbb{R}$. Довести, що статистики $T_1 = \min\limits_{1 \le k \le n} \xi_k$ та $T_2 = \sum\limits_{i=1}^{n} \xi_i - nT_1$ незалежні та мають показникові розподіли, а статистики $T_1 - T_2/n(n-1)$, $T_2/(n-1)$ є незміщеними для b, a. Дослідити їх асимптотику.

6. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з розподілу зі щільністю $f(y, \theta) = \exp\{\theta - y\} \mathbb{1}_{y>\theta}$, а статистика $\xi_{(1)} = \min\limits_{1 \le k \le n} \xi_k$. Довести, що оцінка $\exp(\xi_{(1)} - c)(1 - 1/n) \mathbb{1}_{c > \xi_{(1)}} + \mathbb{1}_{c \le \xi_{(1)}}$ є незміщеною для функції $\tau(\theta) = \mathbf{P}_\theta(\xi_1 > c)$.

7. Нехай ξ_1, ξ_2, ξ_3 – кратна вибірка з рівномірного розподілу на інтервалі $(\theta - 1, \theta + 1)$. Потрібно а) довести, що середнє за величиною з вибіркових значень (медіана) є незміщеною оцінкою для θ; б) для показникового розподілу з середнім θ така оцінка є зміщеною; в) знайти незміщені оцінки для θ як функції від найменшого вибіркового значення.

8. Нехай $\widehat{\theta}_n$ – довільна незміщена консистентна оцінка параметра θ, а подія A_n не залежить від $\widehat{\theta}_n$ і $\mathbf{P}_\theta(A_n) = 1/n$. Довести, що оцінка $\widetilde{\theta}_n = \widehat{\theta}_n \mathbb{1}_{\overline{A_n}} + n^2 \mathbb{1}_{A_n}$ є консистентною, однак вона зміщена та асимптотично зміщена.

9. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з двовимірними квадратично інтегровними спостереженнями $\xi_j = (\xi_{1j}, \xi_{2j})$. Довести, що вибіркова коваріація

$$\widehat{\rho}_n = \frac{1}{n-1} \sum_{j=1}^{n} (\xi_{1j} - \widehat{\mu}_{1n})(\xi_{2j} - \widehat{\mu}_{2n})$$

є незміщеною оцінкою для коваріації $\mathrm{Cov}(\xi_{11}, \xi_{21})$, де $\widehat{\mu}_{in}$ – вибіркове середнє для i-ї координати спостережень.

10. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з розподілу Бернуллі: $\mathbf{P}_\theta(\xi_1 = 1) = \theta = 1 - \mathbf{P}_\theta(\xi_1 = 0)$. Визначити клас параметричних функцій $\tau(\theta)$, для яких існують незміщені оцінки $T(X)$. Довести, що функції: $1/\theta$ та θ^m при $m > n$ не належать до такого класу.

11. Для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ із рівномірного розподілу $U(0, \theta)$ довести, що оцінки $\widehat{\theta}_1 = (n+1)\xi_{(1)}$ та $\widehat{\theta}_2 = (1+\frac{1}{n})\xi_{(n)}$ є незміщеними для θ. Яка з них є кращою ?

12. Побудувати незміщену оцінку характеристичної функції одного спостереження за кратною вибіркою.

13. Розглядається кратна вибірка з двовимірними спостереженнями. Знайти незміщену оцінку для коваріації координат одного спостереження.

14. Для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ із логістичного розподілу зі щільністю $f(x, \theta) = \exp(-x+\theta)(1+\exp(-x+\theta))^{-2}$ незміщеною та консистентною оцінкою параметра θ є вибіркове середнє.

1.1.2. Оцінювання ймовірності успіху у схемі Бернуллі

Література: [1, с. 328–329], [2, с. 331–334]

А

1. В обстеженні, проведеному поштовою компанією, із 200 клієнтів 172 вказали на задоволення часом доставки кореспонденції. Побудувати наближений 95%-й вірогідний інтервал для теоретичної частки задоволених клієнтів.

2. Випадкова вибірка розміру 200 обрана з великої групи страхових полісів. За даними вибірки частка полісів, для яких вимоги

виникли протягом минулого року, становить 0,16. Побудувати наближений 95%-й вірогідний інтервал для істинної частки θ тих полісів, за якими виникли вимоги, серед усіх полісів.

3. Обчислити 99%-й вірогідний інтервал для відсотка вимог певного типу, які повністю врегульовані протягом шести місяців, якщо відомо, що у випадковій вибірці зі 100 вимог цього типу точно 83 були повністю врегульовані протягом шести місяців.

4. Імовірність того, що вимога по певному типу полісу виникне у даному році, дорівнює 0,04. П'ятсот полісів відібрано навмання. Використати нормальне наближення для обчислення ймовірності того, що більше ніж 30 з них призведе до вимоги протягом року.

5. Імовірність того, що інформація в індивідуальному записі бази даних компанії неправильна (наприклад, застаріла), дорівнює 0,13. Обчислити ймовірність того, що у випадковій вибірці з 200 записів не більше 20 містять неправильну інформацію.

6. Спостерігається послідовність з n випробувань Бернуллі з імовірністю успіху θ та відносною частотою $\widehat{\theta}_n$. Довести, що

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \Big/ \left(\widehat{\theta}_n\left(1 - \widehat{\theta}_n\right)\right)^{1/2} \xrightarrow{W} \zeta \simeq N(0,1), n \to \infty.$$

В

1. Під час опитування кожен індивідуум у випадковій вибірці 275 осіб із великої генеральної сукупності повідомляє, яку політичну партію він підтримує. Якщо 45 % генеральної сукупності підтримує партію A, обчислити наближено ймовірність того, що принаймні 116 з опитуваних підтримають A.

2. Великий портфель полісів містить частку $p \in (0,1)$ вимог останнього календарного року. Дослідник вивчає групу з 25 полісів, випадково обраних із портфеля. Використати апроксимацію Пуассона до біноміального розподілу, щоб обчислити наближено ймовірність того, що є принаймні 4 вимоги за останній рік, у кожному з двох випадків: $p = 0,1$ й $p = 0,2$.

3. Для оцінювання ймовірності успіху спостерігається випадкова вибірка X із біноміальним розподілом із параметрами $(20, p)$. Знайти: а) середнє квадратичне відхилення для відносної частоти $\nu_{20}/20$ та обчислити його при $p = 0,5$; б) середнє квадратичне

відхилення для оцінки $(\nu_{20}+1)/(20+1)$, обчислити це значення при $p=0,5$ та порівняти з попереднім.

4. У великій корпорації 50 нових службовців приєдналися до програми пенсійного забезпечення компанії протягом минулого року. Передбачається, що кожен новий службовець має ймовірність 0,40 збереження у схемі протягом принаймні 10 років незалежно для кожного службовця. Обчислити наближене значення ймовірності того, що більше ніж половина 50 торішніх нових службовців залишилися у програмі пенсійного забезпечення протягом принаймні 10 років.

5. Обчислити максимально можливу ширину симетричного двобічного 95%-го вірогідного інтервалу для теоретичної частки населення, що має специфічну характеристику, яка ґрунтується на відповідній інформації у випадковій вибірці розміру 1600.

6. Спостерігається дві незалежні послідовності випробувань Бернуллі з імовірностями успіхів θ_i та кількостями випробувань n_i, $i=1,2$. Нехай $\widehat{\theta}_{n_i}$ – відповідні відносні частоти успіхів, $i=1,2$. Довести, що

$$\sqrt{n_1}\left(\widehat{\theta}_{n_1}-\widehat{\theta}_{n_2}-\theta_1+\theta_2\right)\xrightarrow{W}\zeta\simeq N\left(0,\ \left(\theta_1-\theta_1^2\right)+\rho\left(\theta_2-\theta_2^2\right)\right)$$

при $n_i\to\infty$ так, що $n_1/n_2\to\rho$. Побудувати надійний інтервал для $\theta_1-\theta_2$.

C

1. Журнал стверджує, що 25 % його читачів – студенти. Аналіз випадкової вибірки з 200 читачів засвідчив, що вона містить 42 студенти. Обчислити ймовірність одержання 42 або меншої кількості студентів-читачів, припускаючи, що заява журналу справедлива.

2. Компанія хоче оцінити відсоток її клієнтів, які бажають робити покупки в Інтернеті. Для цього вирішує обчислити симетричний двосторонній 95%-й вірогідний інтервал для невідомого відсотка: а) показати, що на підставі випадкової вибірки з 200 клієнтів необхідний вірогідний інтервал матиме ширину, яка не перевищує 13,9 %; б) обчислити розмір вибірки, який гарантуватиме, що безвідносно до теоретичного відсотка, ширина вірогідного інтервалу не перевищить 10 %.

3. Виникнення вимог у групі з а) 200, б) 2000 полісів моделюється так, що ймовірність вимоги в наступному році дорівнює $0,015$ незалежно для кожного полісу. Кожен поліс може викликати щонайбільше одну вимогу. Обчислити наближене значення ймовірності того, що більше ніж а) 10, б) 40 вимог виникне у цій групі полісів наступного року.

4. Ринкова дослідницька компанія має намір оцінити частку населення θ, що підтримує деяку політичну партію. Вона має намір опитати велику вибірку так, щоб 95%-й вірогідний інтервал для θ мав розмір $0,03$ або менше. Вважають, що θ наближено дорівнює 0,4. Припущено, що кожен, хто потрапив у вибірку, візьме участь в опитуванні. Обчислити мінімальний розмір вибірки, яку компанія повинна використати.

5. Під час екскурсій 84 з 250 чоловіків та 156 з 250 жінок купили сувеніри. Побудувати 95%-й вірогідний інтервал для різниці теоретичних пропорцій чоловіків та жінок, що купили сувеніри.

6. Нехай $\widehat{\theta}_n$ – відносна частота успіху в n випробуваннях Бернуллі з імовірністю успіху θ. Довести, що статистика $n\widehat{\theta}_n\left(1-\widehat{\theta}_n\right)$ є зміщеною і асимптотично незміщеною для дисперсії кількості успіхів.

7. Випадкова величина ζ_θ має розподіл Пуассона: а) довести асимптотичну нормальність $(\zeta_\theta-\theta)/\sqrt{\theta}\xrightarrow{W}\zeta\simeq N(0,1)$; б) для заданого $p\in(0,1)$ побудувати наближений надійний інтервал рівня p для параметра θ з кінцями $\left(\zeta_\theta+x_p^2/2\pm x_p\left(\zeta_\theta+x_p^2/4\right)^{1/2}\right)$.

8. Спостерігається випадкова величина ν, що має біноміальний розподіл із параметром p та невідомою кількістю випробувань n. Знайти надійний інтервал для n.

9. Спостерігається кількість успіхів ν_n у перших n випробуваннях Бернуллі з невідомою ймовірністю успіхів p. Знайти надійний інтервал для кількості успіхів ν_m в наступних m випробуваннях.

1.1.3. Емпірична функція розподілу. Варіаційний ряд. Квантилі

Література: [1, с. 330–341], [2, с. 334–346]

А

1. За вибіркою з неперервного розподілу:

$$2.4,\ 1.0,\ 0.7,\ 0.0,\ 1.1,\ 1.6,\ 1.1,\ -0.4,\ 0.1,\ 0.7$$

записати варіаційний ряд, знайти вибіркову медіану, нижній та верхній квартилі, розмах вибірки. Побудувати графік емпіричної функції розподілу $\widehat{F}_n(x)$ та графіки функцій нормальних розподілів із параметрами $\mu = 0,8$; $\sigma^2 = 0,64$ та $\mu = 1,0$; $\sigma^2 = 0,25$ відповідно. Обчислити значення:

$$\sup_{-\infty<x<+\infty}\left|N_{0,8;0,64}(x) - \widehat{F}_n(x)\right|, \quad \sup_{-\infty<x<+\infty}\left|N_{1,0;0,25}(x) - \widehat{F}_n(x)\right|$$

та показати їх на графіках.

2. Якщо спостереження у кратній вибірці $X = (\xi_1, ..., \xi_n)$ мають щільність f та функцію розподілу F, то щільність порядкової статистики $\xi_{(k)}$ така: $f_k(x) = nC_{n-1}^{k-1}F^{k-1}(x)f(x)(1-F(x))^{n-k}$. Довести це.

3. Знайти розподіл статистик Колмогорова $\widehat{\kappa}_1, \widehat{\kappa}_2$.

4. На графіку "стебла та листя" нижче відображено вартості 40 страхових полісів. Одиниця виміру "стебла" – 10000, одиниця виміру "листя" – 1000.

5	3
5	6
6	02
6	5779
7	122344
7	556677899
8	1123444
8	567778
9	024
9	6

Визначити вибіркові медіану та квартилі для цієї партії полісів.

5. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з нормальним розподілом спостережень $N(\theta, 1)$. Знайти граничний розподіл випадкових величин $\sqrt{n}(\widehat{m}_n - \theta), n \to \infty$, для вибіркової медіани $\widehat{m}_n$. Як співвідносяться асимптотичні ефективності медіани та вибіркового середнього при оцінюванні параметра θ?

В

1. Маємо вибірку з рівномірного розподілу:

$$1.3, 8.6, 3.5, 10.8, 9.3, 7.1, 4.1, 7.9, 2.2, 10.3, 3.2, 8.0, 9.1, 9.6, 6.6.$$

Знайти варіаційний ряд. За цією вибіркою побудувати графік емпіричної функції розподілу $\widehat{F}_n(x)$ та графік функції $F(x)$ рівномірного на [0,11] розподілу. Обчислити $\sup\limits_{-\infty<x<+\infty} \left|F(x) - \widehat{F}_n(x)\right|$.

2. Маємо вибірку: 1.31, 1.18, 1.50, 1.06, 1.01, 1.06, 1.33, 1.80, 1.30, 1.35 з узагальненого показникового розподілу зі щільністю

$$f(y, \theta) = a^{-1} \exp(-(y-b)/a) \mathbb{1}_{y>b}, \quad \theta = (a, b) = (0.5, 1).$$

За цією вибіркою побудувати графік емпіричної функції розподілу $\widehat{F}_n(x)$ та графік показникової функції розподілу $F(x)$ із параметрами $a = 0.5$, $b = 1$. Обчислити $\sup\limits_{-\infty<x<+\infty} \left|F(x) - \widehat{F}_n(x)\right|$.

3. Дані про відсутність на роботі через захворюваність зареєстровано для 30 службовців компанії за 91-денний період. Ці дані наведено нижче:

Кількість відсутніх	0	1	2	3	4	5
Кількість робочих днів	44	19	10	8	7	3

Обчислити вибіркове середнє і стандартне відхилення кількості службовців, що були відсутні протягом хоча б одного дня.

4. Знайти: а) сумісну щільність порядкових статистик $\xi_{(j)}, \xi_{(k)}$, при $j < k$; б) функцію розподілу, математичне сподівання і дисперсію розмаху $\xi_{(n)} - \xi_{(1)}$ для рівномірного розподілу $U([a; b])$.

5. Нехай $(\xi_n, n \geq 1)$ незалежні однаково розподілені величини з неперервною функцією розподілу.

Довести рівність: $\mathbf{P}\left(\xi_n > \max\limits_{1 \leq k < n} \xi_k\right) = 1/n$.

C

1. Довести для функції розподілу Колмогорова рівність

$$K(x) = 1 - 2\sum_{k=1}^{+\infty} (-1)^{k-1} \exp\left(-2k^2x^2\right).$$

2. На відрізку [0,2] навмання обирають точку, яка ділить відрізок на дві частини, і фіксують її координату ξ. Експеримент проведено 10 разів, і при цьому отримано 10 незалежних спостережень випадкової величини ξ:

$$0.00,\ 0.31,\ 0.70,\ 1.40,\ 0.08,\ 1.93,\ 0.79,\ 1.43,\ 1.42,\ 1.69.$$

Знайти варіаційний ряд, вибіркову медіану, нижній та верхній квартилі, розмах вибірки. Побудувати графік емпіричної функції розподілу.

3. Довести, що теоретична медіана $m = x_{1/2}$ є абсолютним центром положення інтегровної випадкової величини ξ, тобто

$$m = \arg\min_{a \in \mathbb{R}} \mathbf{E}\left|\xi - a\right|.$$

4. Довести, що для кратної вибірки X з неперервною функцією розподілу статистика Колмогорова $\widehat{\kappa}_n = \sqrt{n} \sup\limits_{x \in \mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right|$ така:

$$\widehat{\kappa}_n = \sqrt{n} \max_{1 \leq k \leq n} \max\left(\left|k/n - F\left(\xi_{(k)}\right)\right|,\ \left|F\left(\xi_{(k)}\right) - (k-1)/n\right|\right).$$

5. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з неперервною функцією розподілу F. Довести, що статистики $(F(\xi_{(k)})/F(\xi_{(k+1)}))^k$, $1 \leq k < n$, незалежні, та знайти їх розподіл.

6. Нехай $(\xi_1, ..., \xi_n)$ – кратна вибірка з експоненціального розподілу $Exp(\theta)$, а $\left(\xi_{(k)}\right)$ – її варіаційний ряд, $\xi_{(0)} = 0$. Довести, що:
а) сумісна щільність вектора $\left(\xi_{(k)}, k = \overline{1, r}\right)$ дорівнює

$$\theta^r C_n^r r! \exp\left(-\theta\left[\sum_{k=1}^{r} x_k + (n-r)x_r\right]\right);$$

б) випадкова величина $2\theta\left[\sum\limits_{k=1}^{r} \xi_{(k)} + (n-r)\xi_{(r)}\right]$ має хі-квадрат розподіл із $2r$ ступенями вільності;
в) величини $\eta_k = (n-k+1)\left(\xi_{(k)} - \xi_{(k-1)}\right)$, $k = \overline{1, r}$, – незалежні та експоненціально розподілені з параметром θ;
г) у класі лінійних незміщених оцінок для θ^{-1} як функцій від $\xi_{(i)}$, що мають вигляд $\sum\limits_{i=1}^{k} c_i \xi_{(i)}$, найменшу дисперсію має оцінка $k^{-1} \sum\limits_{i=1}^{k} \xi_{(i)} + k^{-1}(n-k)\xi_{(k)}$, а відповідна дисперсія дорівнює $1/(k\theta^2)$.

7. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна вибірка з рівномірним на $[0, 1]$ розподілом спостережень: а) знайти сумісну щільність, математичні сподівання та дисперсії статистик $\xi_{(1)}, \xi_{(n)}$; б) довести, що спейсинги $\delta_k \equiv \xi_{(k+1)} - \xi_{(k)}$, $0 \le k < n$, незалежні, де $\xi_{(0)} = 0$, та знайти їх функції розподілу; в) обчислити функцію розподілу величини $\min\limits_{0 \le k \le n} \delta_k$, де $\xi_{(n+1)} = 1$; г) довести, що $\lim\limits_{n\to\infty} \mathbf{P}(n\xi_{(k)} < x) = \Gamma_{k-1,1}(x)$ для фіксованого k, де $\Gamma_{k\lambda}$ – функція гамма-розподілу з параметрами k, λ.

8. Знайти коваріацію значень емпіричної функції розподілу у двох заданих точках.

9. Довести, що сумісний розподіл значень емпіричної функції розподілу в заданих точках є поліноміальним та знайти його.

10. Нехай $(\xi_n, n \ge 1)$ незалежні однаково розподілені величини з неперервною функцією розподілу, а $\xi_{(n-r+1)}$ – відповідна порядкова статистика для n-кратної вибірки. Розглянемо при $r \ge 1$ випадковий номер $\nu_{n,r} = \min\left(k \ge 1 : \xi_{n+k} \ge \xi_{(n-r+1)}\right)$. Довести: а) $\mathbf{P}(\nu_{n,1} > k) = n/(n+k)$; б) $\mathbf{P}(\nu_{n,r} > k) = C_n^r / C_{n+k}^r$; в) $\lim\limits_{n\to\infty} \mathbf{P}(\nu_{n,r} > nx) = (1+x)^{-r},\ x \ge 0$.

11. Випадкові величини ξ_n, ξ мають однозначно визначені медіани m_n, m. Довести, що з $\xi_n \overset{W}{\to} \xi$, $n \to \infty$, випливає збіжність $m_n \to m$. Навести приклад, коли таке твердження не виконується для математичних сподівань.

12. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з неперервною функцією розподілу F. Знайти граничний розподіл при $n \to \infty$ випадкових величин $n\sqrt{F\left(\xi_{(1)}\right)\left(1 - F\left(\xi_{(n)}\right)\right)}$.

13. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з рівномірним на $[0, 1]$ розподілом. Довести, що

$$\mathbf{P}\left(\widehat{m}_n - 1/2 < x/\sqrt{8n}\right) \to \Phi(x), n \to \infty,$$

де Φ – стандартна нормальна функція розподілу.

14. Обчислити: а) розподіл порядкових статистик, б) сумісний розподіл екстремальних статистик для вибірки із рівномірного розподілу.

15. Нехай $(\xi_n, n \geq 1)$ незалежні рівномірно розподілені на $[0, 1]$ величини. Визначимо випадковий номер

$$\nu = \inf(k \geq 1 : \xi_k < \xi_{k+1}).$$

Довести при $x \in [0, 1], k \geq 1$ тотожність

$$\mathbf{P}(\xi_1 < x, \nu = k) = x^k/k! - x^{k+1}/(k+1)!$$

16. Нехай $\widehat{F}_n(x)$ – емпірична функція розподілу для незалежних спостережень із рівномірним на інтервалі $[0, 1]$ розподілом. Позначимо $p_n(\alpha, \beta) \equiv \mathbf{P}\left(\widehat{F}_n(x) < \alpha + \beta x, \forall x \in [0, 1]\right)$, де $\alpha, \beta > 0$, $\alpha + \beta > 1$. Довести:

а) $p_n(\alpha, \beta) = 1 - ((1 - \alpha)/\beta)^n$ при $\alpha \in [1 - 1/n, 1]$;

б) $p_{n+1}(\alpha, \beta) = \int\limits_{(1-\alpha)/\beta}^{1} p_n\left(\frac{n+1}{n}\alpha, \frac{n+1}{n}\beta t\right)(n+1)t^n dt$ при $\alpha < 1 - 1/n$.

Обчислити $p_n(\alpha, \beta)$.

1.2. Задачі параметричного оцінювання

1.2.1. Вибіркові моменти. Метод моментів

Література: [1, с. 342–348], [2, с. 346–353]

A

1. Нижче вказано кількість гроз, зареєстрованих цим літом 100 метеорологічними станціями:

Кількість гроз	0	1	2	3	4	5
Кількість станцій	22	37	20	13	6	2

Обчислити й порівняти вибіркове середнє і вибіркову медіану кількості гроз.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з пуассонівського розподілу з параметром λ. Оцінити параметр λ за методом моментів.

3. Оцінити значення інтеграла $\int\limits_0^2 (1 + [y]y)^{-1} dy$, де $[y]$ – ціла частина y, за вибіркою: 0.14, 1.81, 1.38, 1.28, 1.39, 0.80, 0.35, 0.70, 1.40, 0.08, 1.93, 0.79, 1.43, 1.42, 1.68 з рівномірного на [0, 2] розподілу. Оцінити дисперсію знайденої оцінки.

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу зі щільністю $f(y, \theta) = a^{-1} \exp\{-(y-b)/a\} \mathbb{1}_{y>b}$, $\theta = (a, b) \in \mathbb{R}_+ \times \mathbb{R}$. Одержати оцінки параметрів a та b методом моментів.

5. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з нормального розподілу $N(\mu_0, \sigma^2)$, μ_0 – відоме. Одержати оцінку для параметра $\theta = \sigma^2$ методом моментів.

6. Довести такі співвідношення між центральними та нецентральними моментами: $\mu_2^0 = \mu_2 - \mu_1^2$, $\mu_3^0 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$, $\mu_n^0 = \sum\limits_{k=0}^{n} C_n^k \mu_{n-k} (-\mu_1)^k$.

B

1. Тривалість роботи елемента до першого виходу з ладу є показниково розподілена випадкова величина. Спостерігалась робота 20 елементів і фіксувалась тривалість роботи до першого виходу з ладу (в годинах): 11, 149, 846, 563, 384, 950, 864, 63, 990, 77,

685, 158, 348, 318, 25, 218, 1803, 63, 1544, 380. Оцінити математичне сподівання тривалості роботи елемента, якщо його міняють, коли тривалість роботи до першого виходу з ладу менша 200 годин, або коли тривалість роботи досягла 200 годин (навіть якщо елемент протягом 200 годин не відмовив). Знайти дисперсію цієї оцінки.

2. Одержати оцінки для параметрів μ та σ нормального розподілу $N(\mu, \sigma^2)$ методом моментів.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з біноміального розподілу з параметрами N і p, де N – відоме. Оцінити параметр p методом моментів.

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з двостороннього показникового розподілу зі щільністю $f(y, \theta) = \exp\{-|y|/\theta\}/2\theta$. Одержати оцінку параметра θ методом моментів.

5. Відношення стандартного відхилення до середнього випадкової величини називають коефіцієнтом варіації. Для кожного з наведених розподілів вирішити, чи при збільшенні середнього збільшиться, зменшиться або не зміниться коефіцієнт варіації: а) Пуассона із середнім λ, б) показниковий із середнім μ, в) хі-квадрат з n ступенями вільності.

6. Знайти щільність, математичне сподівання та дисперсію випадкової величини з логарифмічно-нормальним розподілом.

C

1. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з від'ємно-біноміального розподілу з параметрами r та p (кількість випробувань у схемі Бернуллі з імовірністю успіху p, які необхідно провести до r-го успіху). Одержати оцінки параметрів r та p методом моментів.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з геометричного розподілу з параметром p $(0 < p < 1)$. Оцінити параметр p методом моментів.

3. Довести, що вибіркові моменти та емпірична функція розподілу $\widehat{F}_n$ пов'язані такими співвідношеннями:

$$\widehat{\mu}_{kn} = \int\limits_{-\infty}^{\infty} x^k d\widehat{F}_n(x), \quad \widehat{\mu}_{kn}^0 = \int\limits_{-\infty}^{\infty} (x - \widehat{\mu}_n)^k \, d\widehat{F}_n(x).$$

4. Визначити вибіркові аналоги коефіцієнтів варіації, асиметрії, скошеності та ексцесу як відповідних функцій від вибіркових моментів.

5. Обчислити коефіцієнти асиметрії та ексцесу для стандартних дискретних та абсолютно неперервних розподілів.

6. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з розподілу зі щільністю $f(y, \theta) = 2\theta^{-2}(\theta - y)\mathbb{1}_{0<y<\theta}$, $\theta \in \mathbb{R}_+$. Одержати оцінку параметра θ методом моментів.

7. Для стандартної нормальної випадкової величини $\xi \simeq N(0,1)$ довести інтегруванням за частинами тотожність $\mu_k = (k-1)\mu_{k-2}$ та рівності

$$\mu_{2k} = 2^k \pi^{-1/2}\Gamma(k+1/2) = 1 \cdot 3 \cdots (2k-1) \equiv (2k-1)!!$$

8. Для випадкової величини з гамма-розподілом $\xi \simeq \Gamma(\lambda, \alpha)$ обчислити $\mu_k = \lambda^{-k}\alpha(\alpha+1)\cdots(\alpha+k-1)$.

9. При $n \to \infty$ довести асимптотичне зображення для дисперсії $\mathbf{D}(\widehat{\mu}^0_{kn}) = (\mu_{2k} - \mu_k^2 + k\mu_{k-1}(k\mu_{k-1}\mu_2 - 2\mu_{k+1}))/n \ + \ O(1/n^2)$.

10. Довести, що вибіркове середнє для кратної вибірки із щільністю Коші $1/\pi(1+(y-\theta)^2)$ не є консистентною оцінкою центра симетрії θ, оскільки розподіл статистики не залежить від n.

11. Нехай $(\xi_1, \ldots, \xi_n)$ – кратна вибірка з неперервною функцією розподілу F, варіаційним рядом $\left(\xi_{(1)}, \ldots, \xi_{(n)}\right)$, а $\alpha \in (0, 1/2)$. Для усунення впливу можливих випадкових збурень із занадто великими чи малими значеннями використовують робасний (стійкий до збурень) аналог вибіркового середнього:

$$\widehat{\mu}_{n\alpha} \equiv (n(1-2\alpha))^{-1} \sum_{k=[n\alpha]+1}^{[n-n\alpha]} \xi_{(k)}.$$

Довести: а) $\widehat{\mu}_{n,0} = \widehat{\mu}_n$; б) $\widehat{\mu}_{n\alpha} \overset{P1}{\to} \int\limits_{x_\alpha}^{x_{1-\alpha}} y dF(y)$, $n \to \infty$, де x_p – квантиль рівня p для F.

12. Для покращення властивостей оцінок пропонується метод розщеплень (bootstrap). Для довільних оцінок $T_n = T_n(\xi_1, \ldots, \xi_n)$ обирають оцінку вигляду $T_n^* = \frac{1}{n}\sum\limits_{k=1}^{n} T_{n-1}(\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_n)$. Довести, що застосування методу розщеплень до статистик $\widehat{\sigma}^2_n$ при $n > 2$ приводить до статистик $\widehat{s}^2_n$.

13. Для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ з $\mu_4 < \infty$ довести асимптотичну нормальність:

$$\sqrt{n}\left(\widehat{s}_n^2 - \sigma^2\right) \xrightarrow{W} N\left(0, \mu_4^0 - \sigma^4\right), n \to \infty.$$

14. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з показниковим розподілом $Exp(\theta)$. Знайти оцінку методу моментів для θ у вигляді функції від а) першого вибіркового моменту, б) другого вибіркового моменту, в) перших двох вибіркових моментів; а також г) знайти таку оцінку для ймовірності $\mathbf{P}_\theta(\xi_1 \geq 1)$.

15. Нехай F_n – функція розподілу з моментами $(\mu_k(n), k \geq 0)$, і при кожному k існує границя $m_k = \lim\limits_{n\to\infty} \mu_k(n)$.

Довести: а) послідовність $(m_k, k \geq 0)$ утворена моментами деякої функції розподілу F; б) якщо F – нормальна функція розподілу, то $F_n \xrightarrow{W} F, n \to \infty$.

16. Нехай випадкова величина ξ набуває значень на відрізку $[0,1]$: а) довести, що послідовність нецентральних моментів $(\mu_n, n \geq 0)$ з $\mu_0 \equiv 1$ є цілком монотонною, тобто $(-1)^k \Delta^k \mu_n \geq 0$ для всіх $n \geq 0, k \geq 1$, де різницевий оператор визначається через $\Delta x_n \equiv x_{n+1} - x_n, n \geq 0$; б) вивести, що $\sum\limits_{k<nx} C_n^k (-1)^{n-k} \Delta^{n-k} \mu_k \xrightarrow{O} \mathbf{P}(\xi < x),\ n \to \infty$, за теоремою Бернштейна; в) застосувати формулу б) до послідовностей $\mu_n = p^n$, $1/(n+1)$, $2/(n+2)$.

17. Довести асимптотичну нормальність вибіркової дисперсії.

18. Знайти оцінку методу моментів для кратної вибірки з подвійним розподілом Пуассона:

$$\mathbf{P}(\xi_1 = k) = (\exp(-\theta_1)\theta_1^n + \exp(-\theta_2)\theta_2^n)/2n!$$

19. Нехай $\widehat{\theta}_n(X)$ – оцінка параметра θ за кратною вибіркою $X = (\xi_1, \ldots, \xi_n)$. Оцінка дисперсії $\widehat{\theta}_n(X)$ методом складаного ножа (jackknife) має вигляд

$$\widehat{\sigma}_n^2 = \frac{n-1}{n} \sum_{i=1}^n \left(\widehat{\theta}_{ni}(X) - n^{-1} \sum_{j=1}^n \widehat{\theta}_{nj}(X)\right)^2,$$

де $\widehat{\theta}_{ni}(X) = \widehat{\theta}_{n-1}(\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n)$. Нехай $\widehat{\theta}_n(X) = (\widehat{\mu}_n)^2$. Довести, що

$$\widehat{\sigma}_n^2 = 4(\widehat{\mu}_n)^2\widehat{\mu}_{2n}^0/(n-1) - 4\widehat{\mu}_n\widehat{\mu}_{3n}^0/(n-1)^2 + \left(\widehat{\mu}_{4n}^0 - \left(\widehat{\mu}_{2n}^0\right)^2\right)/(n-1)^3.$$

1.2.2. Незміщені оцінки. Нерівність Крамера – Рао

Література: [1, с. 349–361], [2, с. 353–367]

A

1. Знайти незміщену оцінку для параметра σ за кратною вибіркою об'єму 2 з нормального розподілу $N(a, \sigma^2)$ із невідомим a.

2. Знайти оптимальну оцінку для параметра σ^2 за кратною вибіркою з нормального розподілу $N(a, \sigma^2)$, де a – відоме середнє (*критерій оптимальності*).

3. Нехай X – кратна вибірка з гамма-розподілу $\Gamma(\lambda, \alpha)$. Довести, що вибіркове середнє $\widehat{\mu}_n$ є оптимальною незміщеною оцінкою для α/λ.

4. У схемі Бернуллі розглянемо клас статистик вигляду

$$T_{\alpha\beta} = (\nu_n(X) + \alpha)/(n + \beta), \;\; \alpha, \beta \in \mathbb{R};$$

а) обчислити квадратичне відхилення $T_{\alpha\beta}$ від θ; б) довести, що при $\alpha = \sqrt{n}/2$ та $\beta = \sqrt{n}$ це відхилення не залежить від θ і дорівнює $(2\sqrt{n}+2)^{-2}$; в) порівняти вказане значення з таким самим відхиленням оптимальної оцінки T^*.

5. Чи є ефективною оцінка за методом максимальної вірогідності $\widehat{\theta} = n\left(\sum\limits_{i=1}^{n} \xi_i\right)^{-1}$ параметра θ показникового розподілу зі щільністю $f(y, \theta) = \theta e^{-\theta y}$, $y \geq 0$?

B

1. Знайти незміщену оцінку для параметра σ за кратною вибіркою об'єму 2 з нормального розподілу $N(0, \sigma^2)$.

2. Знайти оптимальну оцінку для середнього a за кратною вибіркою з нормального розподілу $N(a, \sigma^2)$, де a та σ^2 – невідомі параметри (*критерій оптимальності для векторного параметра*).

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з рівномірного розподілу на $(0, \theta)$, $\theta \in (0, 1)$, $\xi_k \leq \theta$ м.н. Знайти незміщену оцінку параметра θ та дослідити її на ефективність (*приклад суперефективної оцінки.*)

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з пуассонівського розподілу з параметром $\lambda > 0$. Чи є оцінка $\lambda = n^{-1} \sum\limits_{i=1}^{n} \xi_i$ ефективною оцінкою параметра λ? Довести, що для значення функції $g(\lambda) = 1/\lambda$ не існує незміщеної оцінки.

5. Довести, що розподіли: біноміальний, Пуассона, негативний біноміальний, експоненціальний, нормальний, багатовимірний нормальний є окремими випадками експоненціальної моделі.

C

1. Довести, що при $|\Theta| > 1$ у класі всіх оцінок параметра θ не існує такої, що має рівномірно найменше середньоквадратичне відхилення.

2. Довести, що в умовах теореми Крамера – Рао для векторного параметра матриця $\mathrm{Cov}_\theta(T(X) - \tau(\theta)) - \tau'_\theta I^{-1}(\theta)\tau_\theta$ для векторної функції $\tau(\theta)$ та для її незміщеної оцінки $T = T(X) \in \Gamma_\tau$ є невід'ємно визначеною, зокрема, має невід'ємні діагональні елементи.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з біноміального розподілу

$$P\left(k, \theta\right) = C_N^k \theta^k \left(1 - \theta\right)^{N-k}, \;\; k = \overline{0, N},$$

де N – відоме ціле число, θ належить деякому скінченому проміжку $[a, b], a > 0, b < 1$. Чи є оцінка $\theta = \frac{1}{nN} \sum\limits_{i=1}^{n} \xi_i$ ефективною оцінкою параметра θ?

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з двостороннього показникового розподілу зі щільністю $f\left(y, \theta\right) = \frac{1}{2\theta} \exp\left\{-\frac{1}{\theta}\left|y\right|\right\}$, де θ належить деякому скінченному проміжку $[a, b]$, $a > 0$. Чи є оцінка $\theta = \frac{1}{n} \sum\limits_{i=1}^{n} |\xi_i|$ ефективною оцінкою параметра θ ?

5. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з показникового розподілу зі щільністю $f\left(y, \theta\right) = \theta^{-1} \exp\{\, -y/\theta\} \mathbb{1}_{y>0}$, де θ належить

деякому скінченному проміжку $[a, b]$, $a < b$. Чи є оцінка $\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ ефективною оцінкою параметра θ?

6. Нехай X_1, X_2 – незалежні кратні вибірки об'ємів n_1, n_2 з нормальних розподілів $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ відповідно, а $\widehat{\mu}_1, \widehat{\mu}_2$ – вибіркові середні для цих вибірок.

Довести: а) при $\alpha \in [0,1]$ оцінка $\widehat{\mu}_\alpha = \alpha\widehat{\mu}_1 + (1-\alpha)\widehat{\mu}_2$ є незміщеною для μ; б) її дисперсія найменша при $\alpha = \frac{n_1\sigma_1^{-2}}{n_1\sigma_1^{-2}+n_2\sigma_2^{-2}}$.

7. Нехай X – n-кратна вибірка з логістичною щільністю спостережень $f(y,\theta) = \exp(\theta - y)(1+\exp(\theta-y))^{-2}, y \in \mathbb{R}, \theta \in \mathbb{R}$. Довести: а) вибіркове середнє $\widehat{\mu}_n$ є незміщеною оцінкою для θ з дисперсією $\pi^2/3n$, б) функція інформації дорівнює $I_n(\theta) = n/3$, тобто вибіркове середнє у цьому разі не є ефективною оцінкою, хоча має непогану ефективність $9/\pi^2$.

8. За кратною вибіркою $X = (\xi_1, \ldots, \xi_n)$ із рівномірного розподілу $U(\theta, 2\theta)$ знайти оптимальну незміщену оцінку у класі лінійних оцінок вигляду $\widehat{\theta} = \alpha\xi_{(1)} + (1-\alpha)\xi_{(n)}, \alpha \in [0,1]$.

9. Довести, що оптимальна оцінка $T = T(X)$ за кратною вибіркою $X = (\xi_1, \ldots, \xi_n)$ параметра θ є симетричною функцією від X. Для цього довести, що дисперсія симетризованої оцінки не перевищує дисперсії вихідної незміщеної оцінки.

10. Довести, що для кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ із логістичним розподілом спостережень

$$f(y,\theta) = \exp(\theta - y)(1+\exp(\theta - y))^{-2},\ y \in \mathbb{R},$$

вибіркове середнє $\widehat{\mu}_n$ є незміщеною та консистентною оцінкою для параметра θ, однак не є ефективною оцінкою.

11. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка із біноміальним розподілом спостережень із параметрами k, θ. Описати параметричні функції $\tau(\theta)$, для яких існує незміщена оцінка вигляду $g(\xi_1 + \ldots + \xi_n)$ із борелівською функцією g.

12. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з нормальним розподілом $N(0, \theta^2)$. Довести, що статистика $\widehat{\theta} = \sqrt{\pi/2}\, n^{-1}\sum\limits_{j=1}^{n}|\xi_j|$ є незміщеною та консистентною оцінкою параметра θ.

13. Нехай $X = (\xi_1, \dots, \xi_n)$ – кратна вибірка з гамма-розподілом $\Gamma(\theta, \alpha)$ спостережень, а функція $\tau(\theta) = \theta^{-2}$. Довести, що її оцінка вигляду $\widehat{\tau} = n(\widehat{\mu}_n)^2/\alpha(n\alpha + 1)$ є оптимальною незміщеною для τ та не є ефективною.

14. Кратна вибірка $X = (\xi_1, \dots, \xi_n)$ містить спостереження зі щільністю $\theta f_1(x) + (1-\theta) f_2(x)$ із відомими різними щільностями $f_i(x)$. Знайти оптимальну оцінку для параметра θ.

15. Нехай функція впливу $U(X, \theta)$ центрована: $\mathbf{E}_\theta U(X, \theta) = 0$. Довести, що наступні твердження щодо статистики η еквівалентні: а) розподіл η не залежить від θ, б) $\mathrm{Cov}_\theta(\eta, U(X, \theta)) = 0$, в) $\mathrm{Cov}_\theta(\exp(it\eta), U(X, \theta)) = 0,\ t \in \mathbb{R}$.

16. Спостерігається кратна вибірка

$$X = (\max(\xi_1, a), \dots, \max(\xi_n, a)),$$

де величини ξ_j незалежні та однаково розподілені з розподілом Пуассона $\Pi(\theta)$. Знайти незміщену оцінку для θ та її дисперсію.

1.2.3. Достатні статистики та оптимальність

Література: [1, с. 362–366], [2, с. 367–372]

А

1. Знайти функцію вірогідності та достатні статистики для кратної вибірки $\xi_1, \dots, \xi_n$: а) з рівномірного розподілу $U(a, b)$, б) узагальненого показникового розподілу зі щільністю $f\,(y, \theta) = a^{-1} \exp\{-(y-b)/a\}\, \mathbb{1}_{y>b},\ \theta = (a, b) \in \mathbb{R}_+ \times \mathbb{R}$.

2. Для пуассонівської кратної вибірки з невідомим параметром θ вибіркова сума $\widehat{S}_n$ є повною достатньою статистикою. Знайти умовний розподіл вибірки за умови $\widehat{S}_n = s$. Зокрема, оптимальною оцінкою для полінома $\sum\limits_{k=1}^{m} c_k \theta^k$ є статистика $\sum_{k=1}^{m} c_k \pi_k(\widehat{S}_n) n^{-k}$, де $\pi_k(s) = s(s-1)\cdots(s-k+1)$.

3. Якщо T – достатня статистика, а φ – вимірна бієкція, то $\varphi(T)$ – також достатня статистика. Довести це.

4. Довести, що для кратної вибірки $X = (\xi_1, ..., \xi_n)$ із гамма-розподілом $\xi_k \simeq \Gamma(\lambda, \alpha)$ пара статистик $\left(\sum\limits_{k=1}^{n} \xi_k, \sum\limits_{k=1}^{n} \ln \xi_k\right)$ є достатньою статистикою. Враховуючи (без доведення), що ця статистика

є повною, знайти оптимальні оцінки для а) α/λ, б) $(\lambda/(\alpha-1))^n$, в) $n\alpha(n\alpha+1)\cdots(n\alpha+k-1)$ при заданих k та $\lambda=1$.

5. Для кратної вибірки X із неперервною функцією розподілу спостережень її варіаційний ряд є достатньою статистикою при будь-якій параметризації розподілу спостереження.

В

1. Знайти функцію вірогідності та достатні статистики для кратної вибірки $\xi_1,\ldots,\xi_n$ а) з біноміального розподілу, б) нормального розподілу $N(\mu,\sigma^2)$.

2. Нехай $X=(\xi_1,\ldots,\xi_n)$ – кратна вибірка з експоненціального розподілу $Exp(\theta)$. Довести, що оптимальна незміщена оцінка для функції $\exp(-\theta a)$ дорівнює $(\max(0,1-a/S_n))^{n-1}$, де $S_n=\xi_1+ +\ldots+\xi_n$, $a\geq 0$.

3. Нехай $X=(\xi_1,\ldots,\xi_n)$ – кратна вибірка з дискретним розподілом спостережень степеневого вигляду:

$$\mathbf{P}_\theta(\xi_1=y)=a(y)\theta^y/f(\theta), y\in\mathbb{Z}_+,$$

де $\theta\in\Theta=(0,r)\subset\mathbb{R}_+$, причому $f(\theta)=\sum_{y\geq 0}a(y)\theta^y<\infty$ для всіх $\theta<r$. Довести: а) статистика $T=\sum_{k=1}^{n}\xi_k$ є достатньою, б) розподіл T має аналогічний степеневий вигляд, в) статистика T є повною; а також г) знайти оптимальну оцінку для функції $\tau(\theta)=\theta^k$.

4. Вибірка $X=(\xi_1,...,\xi_n)$ містить незалежні спостереження $\xi_j\simeq N(\alpha+\beta c_j,\sigma^2)$, де c_j – відомі коефіцієнти. Довести, що статистика $\left(\sum_{j=1}^{n}\xi_j,\ \sum_{j=1}^{n}c_j\xi_j,\ \sum_{j=1}^{n}\xi_j^2\right)$ є достатньою.

5. Для кратної вибірки $X=(\xi_1,...,\xi_n)$ із поліноміальним розподілом $\mathbf{P}(\xi_1=x_i)=p_i, i=\overline{1,m}$, $\sum p_i=1$, довести, що статистика $\nu(n)=(\nu_{in}, i=\overline{1,m})$ є повною, де $\nu_{in}=\sum_{j=1}^{n}\mathbb{1}_{\{\xi_j=x_i\}}$ – частота значення x_i. Чи є ця статистика мінімальною ?

C

1. Для кратної нормальної вибірки з невідомими середнім та дисперсією (μ, σ^2): а) довести, що вибіркові середнє та дисперсія утворюють повну достатню статистику, б) знайти оптимальну оцінку для $\Phi(-\mu/\sigma)$, де Φ – стандартна нормальна функція розподілу.

2. Вибірка утворена випадковою величиною X із дискретним розподілом $\mathbf{P}(X = -1) = \theta$, $\mathbf{P}(X = k) = (1-\theta)^2\theta^k$, $k = 0, 1, \ldots$, а параметр $\theta \in (0, 1)$. Довести: а) X – достатня та неповна статистика, б) оцінка $1\!\!1_{\{X=0\}}$ – оптимальна для $(1-\theta)^2$.

3. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з експоненціального розподілу $Exp(\theta)$. Довести: а) оптимальна незміщена оцінка для значення функції $\sum\limits_{k=0}^{m} c_k \theta^k / k!$ при $m < n$ дорівнює $\sum\limits_{k=0}^{m} c_k C_{n-1}^k S_n^{-k}$, де $S_n = \xi_1 + \ldots + \xi_n$; б) при $m = n$ така оцінка не завжди існує.

4. Для n-кратної нормальної вибірки з невідомими середнім θ та одиничною дисперсією довести, що оптимальна оцінка для $\theta^k, k \in \mathbb{N}$, дорівнює $(-1)^k n^{-k/2} H_k(-\sqrt{n}\widehat{\mu}_n)$, де за означенням поліноми Ерміта дорівнюють

$$H_k(x) \equiv (-1)^k \exp(x^2/2)\ d^k/dx^k \exp(-x^2/2).$$

5. Для n-кратної вибірки: а) з геометричним розподілом $G(\theta)$, б) показниковим розподілом $Exp(\theta)$ вибіркова сума $\widehat{S}_n$ є повною достатньою статистикою. Знайти умовний розподіл вибірки за умови $\widehat{S}_n = s$. У випадку б) оптимальна оцінка для функції $\exp(-\theta a)$ дорівнює $\left(1 - \min(a, \widehat{S}_n)/\widehat{S}_n\right)^n$.

6. Довести, що для кратної вибірки X з розподілу Коші зі щільністю спостережень $f(y) = (\pi(1 + (y-\theta)^2))^{-1}$ єдиною достатньою статистикою є X.

7. Довести, що статистика S є мінімальною достатньою статистикою для вибірки X тоді і тільки тоді, коли

$$\sigma[S] = \cap_{\{T \text{ – достатня}\}} \sigma[T].$$

Вивести звідси існування мінімальної достатньої статистики.

8. Довести, що статистика S є мінімальною достатньою статистикою для вибірки X, якщо для довільної достатньої статистики T статистика S є достатньою для T. Знайти мінімальну достатню статистику: а) у схемі Бернуллі, б) для вибірки зі щільністю $2(1-y/\theta)^+$, в) для вибірки зі щільністю $\exp(-y+\theta-\exp(-y+\theta))$.

9. Нехай $(\xi_1,\ldots,\xi_n)$ – кратна вибірка з рівномірного розподілу на інтервалі $(-\theta,\theta)$. Довести, що $\widehat{\theta} = \max\limits_{1\le k\le n} \max(\xi_k,-\xi_k)(1+1/n)$ є оптимальною оцінкою параметра θ.

10. Нехай $(\xi_1,\ldots,\xi_n)$ – кратна вибірка з рівномірного розподілу на інтервалі $(\theta,\theta+1)$. Довести, що одновимірна достатня статистика не існує, а двовимірна статистика $\left(\xi_{(1)},\xi_{(n)}\right)$ є достатньою та неповною.

11. Якщо T – повна достатня статистика, а розподіл статистики S не залежить від параметра θ, то випадкові величини T і S – незалежні.

12. Для кратної вибірки X зі щільністю спостережень

$$f(y,\theta) = (\mathbb{1}_{y\in[0,\theta)} + 2\cdot \mathbb{1}_{y\in[\theta,2\theta)})/3\theta$$

існує одновимірна достатня статистика. Знайти її.

13. Довести, що ефективна оцінка скалярної функції $\tau(\theta)$ є достатньою.

14. Знайти оптимальну незміщену оцінку для площі кола через кратну вибірку вимірювань його радіуса у припущенні нормальності та відсутності систематичності похибок.

15. Елементи кратної вибірки набувають значень з інтервалу $[a(\theta),b]$ із відомою монотонною функцією $a(\theta)$. Довести, що екстремальна статистика $\xi_{(1)}$ є достатньою тоді і тільки тоді, коли щільність спостереження має вигляд $f(y,\theta) = g(y)/h(\theta)\mathbb{1}_{y\in[a(\theta),b]}$.

16. Якщо статистика η не залежить від достатньої статистики, то розподіл η не залежить від θ.

1.2.4. Оцінки максимальної вірогідності-1

Література: [1, с. 367–376], [2, с. 372–383]

A

1. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з пуассонівського розподілу з параметром $\lambda > 0$. Оцінити параметр λ методом максимальної вірогідності та довести консистентність оцінки (враховуючи закон великих чисел).

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з від'ємно-біноміального розподілу з відомим параметром $r \geq 1$ та невідомою ймовірністю успіху p. Оцінити параметр p методом максимальної вірогідності та довести консистентність оцінки.

3. У тесті контролю якості відібрано випадкову вибірку зі 100 виробів, якість 90 з яких визнана задовільною. Імовірність p того, що виріб має задовільну якість, є невідомою. Обчислити ймовірність спостереження 90 виробів задовільної якості зі 100 як функцію від p, і в такий спосіб одержати оцінку максимальної вірогідності для p. Знайти відповідний вірогідний інтервал з урахуванням асимптотичної нормальності оцінки максимальної вірогідності (ОМВ).

4. Популяція бегемотів містить невідому кількість N особин. Випадково обрані n з них були помічені та відпущені. З числа m наступних випадково відловлених особин виявилось k помічених. Знайти оцінку максимальної вірогідності для N на підставі вказаної інформації.

5. На початку кожного з 20 холодних днів березня фермер намагається завести трактор. Для успішного заведення йому знадобилися такі кількості спроб: 1, 3, 5, 1, 2, 1, 3, 7, 2, 4, 4, 8, 1, 3, 6, 5, 2, 1, 2, 4. Виходячи з моделі геометричного розподілу числа спроб, знайти ОМВ його параметра. Знайти відповідний вірогідний інтервал з урахуванням асимптотичної нормальності ОМВ.

B

1. Знайти функцію вірогідності, достатні статистики та оцінки максимальної вірогідності для вибірки $\xi_1, \ldots, \xi_n$ з геометричного розподілу. Довести консистентність оцінки.

2. Випадкова величина ξ має розподіл Пуассона з параметром θ. За умови, що нульового значення вона не набуває: а) довести, що розподіл ξ має вигляд $p(y) = \theta^y \exp(-\theta)/(y!(1-\exp(-\theta)))$, $y = 1, 2, \ldots$; б) обчислити $\mathbf{E}Y = \theta/(1-\exp(-\theta))$; в) нехай $(\xi_1, ...\xi_n)$ – вибірка з розподілу p. Довести, що оцінка максимальної вірогідності $\widehat{\theta}$ є розв'язком рівняння $\widehat{\mu}_n - \widehat{\theta} \ / \ (1 - \exp(-\widehat{\theta})) = 0$, та довести, що оцінка максимальної вірогідності збігається з оцінкою методу моментів; г) отримати вираз для нижньої межі Крамера – Рао дисперсії незміщеної оцінки θ.

3. Знайти ОМВ для невідомої кількості спостережень n за біноміальною одноелементною вибіркою $X \simeq B(n, p)$ при відомому p.

4. Для групи страхових полісів кількість вимог для кожного полісу моделюється розподілом Пуассона з інтенсивністю λ на рік незалежно для кожного полісу. Така група з 1500 полісів викликала 183 вимоги протягом минулого року: а) знайти оцінку максимальної вірогідності для λ; б) звідси отримати оцінки величин (1) – імовірність того, що підмножина з 10 полісів не матиме жодних вимог за наступні шість місяців, (2) – імовірність того, що набір із 250 із цих полісів викличе більше ніж 40 вимог за наступний рік.

5. Обчислити інформацію за Кульбаком для кратних вибірок із біноміальним розподілом при відомому об'ємі n та невідомому параметрі $\theta = p$.

C

1. Нехай функція вірогідності $L(X, \theta)$ неперервна за θ, а ОМВ $\widehat{\theta}_n$ визначена однозначно та є достатньою статистикою. Довести, що $\widehat{\theta}_n$ – мінімальна достатня статистика.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з біноміального розподілу з параметром $p \in (0, 1)$. Оцінити параметр p методом максимальної вірогідності.

3. Кількість відвідувань S_n певного сайта за перші n днів має розподіл Пуассона $\Pi(n\lambda)$. За наступні m днів сайт відвідали S_{1m} вітчизняних та S_{2m} зарубіжних користувачів, ці кількості незалежні та мають розподіли Пуассона $\Pi(m\lambda_1)$ та $\Pi(m\lambda_2)$ відповідно,

причому $\lambda = \lambda_1 + \lambda_2$. Знайти ОМВ для (λ_1, λ_2) на підставі спостережень S_n, S_{1m}, S_{2m}.

4. Нехай X, Y – дві незалежні вибірки однакового об'єму n із нормальними розподілами спостережень $N(\alpha + \beta, 1)$, $N(\alpha - \beta, 1)$ відповідно. Знайти ОМВ параметрів α, β.

5. Нехай $X = (\xi_1, \ldots, \xi_n), Y = (\zeta_1, \ldots, \zeta_m)$ – незалежні кратні вибірки з нормальними розподілами $N(\mu_1, \sigma^2)$ та $N(\mu_2, \sigma^2)$ відповідно. Знайти ОМВ для параметра $\theta = (\mu_1, \mu_2, \sigma^2)$.

6. Спостерігаються нормальні випадкові величини $\xi_k \simeq N(\theta_k, 1)$, $k = 1, 2$. Знайти: а) ОМВ для $\theta = (\theta_1, \theta_2)$, б) вказану ОМВ у припущенні, що $\theta_1 \le \theta_2$.

7. Нехай $(\xi_1, \ldots, \xi_n)$ – кратна вибірка розміру n із розподілу зі щільністю $f(y, \theta) = \alpha\varphi((y - \mu)/\sigma)/\sigma + (1 - \alpha)\varphi(y - \mu)$, де φ – стандартна нормальна щільність розподілу. Довести, що ОМВ для параметра $\theta = (\mu, \sigma^2)$ не існує.

8. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з нормальним розподілом на площині: $\xi_1 \simeq N_2(0, 0, 1, 1, \rho)$: а) знайти ОМВ для ρ та обчислити її асимптотичну дисперсію; б) перевірити незміщеність вибіркового коефіцієнта кореляції $\widehat{\rho}_n = \frac{1}{n} \sum_{j=1}^{n} \xi_{1j}\xi_{2j}$ та знайти його асимптотичну дисперсію.

9. Знайти оцінку максимальної вірогідності для параметра μ за кратною вибіркою з розподілу $N(\mu, \sigma^2)$ за умови, що $\mu \in \mathbb{Z}$.

10. Обчислити оцінку максимальної вірогідності для параметра θ за кратною вибіркою з розподілу $N(\theta, 2\theta^2)$. Довести її консистентність.

11. Записати рівняння максимальної вірогідності для кратної вибірки з нормальним розподілом $N(\theta, v(\theta))$ при відомій функції v. Розглянути випадок $v(\theta) = \theta^k$.

12. Спостерігається кратна вибірка об'єму n із нормального розподілу із середнім μ та дисперсією, що для кожного спостереження незалежно від інших дорівнює 1 чи σ^2 з імовірностями $1/2$. Тут невідомими параметрами є μ та σ^2. Знайти ОМВ та довести, що не існує їх границі за ймовірністю при $n \to \infty$.

13. Довести, що середнє квадратичне відхилення оцінки максимальної вірогідності $\widehat{\sigma}^2_n$ від параметра σ^2 для кратної вибірки

з розподілом $N(\mu, \sigma^2)$ та невідомими μ, σ^2 строго більше за таке саме відхилення оцінки $n\widehat{\sigma}_n^2/(n+1)$. Порівняти його з дисперсією незміщеної оцінки $\widehat{s}_n^2$.

14. Виписати рівняння максимальної вірогідності для коефіцієнта кореляції θ для кратної вибірки із двовимірних спостережень із нормальним розподілом $N_2(0,0,1,1,\theta)$.

1.2.5. Оцінки максимальної вірогідності-2

Література: [1, с. 367–376], [2, с. 372–383]

А

1. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з рівномірного розподілу на інтервалі $(-\theta, \theta)$, де θ – невідомий додатний параметр. Вибірка X розміру 5 дає значення $0.87, -0.43, 0.12, -0.92$, і 0.58: а) зобразити наближений графік функції вірогідності $L(X, \theta)$ проти θ для цієї вибірки; б) знайти оцінку максимальної вірогідності; в) чи виглядає ця оцінка прийнятною з огляду на малу кількість спостережень?

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з показникового розподілу зі щільністю $f(y, \theta) = a^{-1} \exp\{-(y-b)/a\} \mathbb{1}_{y>b}$, де $\theta = (a, b) \in \in \mathbb{R}_+ \times \mathbb{R}$. Одержати оцінки параметрів a та b методом максимальної вірогідності.

3. Кратна вибірка X утворена спостереженнями з нормальним розподілом $N(\theta, 2\theta)$. Довести, що ОМВ параметра θ має вигляд $\sqrt{1 + \sum\limits_{k=1}^{n} \xi_k^2 \;/n} - 1$ і є конзистентною.

4. Відсоток відшкодування на інвестиціях певного типу протягом одного року моделюється як нормальна випадкова величина ξ із середнім μ і дисперсією 1. Потенційний інвестор цікавиться таким шансом: вказаний відсоток на таких інвестиціях перевищить 9 %. Випадкова вибірка з десяти таких відсотків дорівнює 7.3, 8.9, 8.3, 6.2, 9.8, 7.7, 9.4, 7.9, 9.1, 7.4: а) обчислити оцінку максимальної вірогідності для $\theta = \mathbf{P}(\xi > 9)$; б) знайти відповідний вірогідний інтервал з урахуванням асимптотичної нормальності ОМВ.

5. Довести, що ОМВ центра симетрії θ розподілу зі щільністю Лапласа $(1/2)\exp(-|y-\theta|)$ збігається з вибірковою медіаною. Порівняти її дисперсію з дисперсією вибіркового середнього.

В

1. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з рівномірного розподілу $U(a,b)$: а) оцінити параметри a та b за методом максимальної вірогідності; б) знайти розподіл оцінок та побудувати вірогідний інтервал для a.

2. Знайти функцію вірогідності, достатні статистики та оцінки максимальної вірогідності для вибірки $\xi_1, \ldots, \xi_n$:
а) з логарифмічно-нормального розподілу зі щільністю

$$f\left(y, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\}, \quad y > 0,$$

б) з розподілу Релея зі щільністю

$$f(y, \theta) = \begin{cases} \frac{y}{\theta} \exp\left\{-\frac{y^2}{2\theta}\right\}, & y > 0, \\ 0, & y \leq 0, \end{cases} \qquad \theta > 0.$$

3. Знайти оцінку максимальної вірогідності для кратної вибірки зі щільністю спостережень $\exp(-y+\theta-\exp(-y+\theta))$, $y, \theta \in \mathbb{R}$.

4. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з двостороннього показникового розподілу зі щільністю

$$f(y, a, b) = (2a)^{-1} \exp\left\{-\left|y-b\right|/a\right\}, \quad a > 0.$$

Одержати оцінки параметрів a та b методом максимальної вірогідності.

5. Обчислити інформацію за Кульбаком для кратних вибірок: а) із нормальним розподілом $N(\mu, \sigma^2)$; б) із гамма-розподілом $\Gamma(\lambda, \alpha)$ при невідомих параметрах.

C

1. За вибіркою $\xi_1, \ldots, \xi_n$ з розподілу зі щільністю спостережень $f(y,\theta) = \theta \exp\{-\theta y\} \mathbb{1}_{y>0}$ одержати оцінку максимальної вірогідності параметра θ.

2. Нехай $(\xi_1, \ldots, \xi_n)$ – кратна вибірка розміру n із розподілу зі щільністю $f(y,\theta) = 2y/\theta^2$, $0 < y < \theta$: а) записати функцію вірогідності та довести, що оцінка максимальної вірогідності для θ така: $\widehat{\theta} = \max(\xi_1, \ldots, \xi_n)$; б) без обчислення $\widehat{\theta}$ пояснити, чому $\widehat{\theta}$ – зміщена оцінка θ.

3. Нехай функція вірогідності $L(X,\theta)$ неперервна за θ, а ОМВ $\widehat{\theta}_n$ визначена однозначно та є достатньою статистикою. Довести, що $\widehat{\theta}_n$ – мінімальна достатня статистика.

4. Довести, що ОМВ для вибірки з рівномірним розподілом спостережень $U(\theta, 1+\theta)$ утворюють цілий інтервал.

5. Знайти ОМВ функції $\tau(\theta) = \Phi(-\mu/\sigma) \equiv \mathbf{P}_\theta(\xi_1 < 0)$ для нормальної кратної вибірки $X = (\xi_1, \ldots, \xi_n)$, $\xi_k \simeq N(\mu, \sigma^2)$ із невідомим параметром $\theta = (\mu, \sigma^2)$. Знайти асимптотичний розподіл для цієї оцінки після її центрування на $\tau(\theta)$ та нормування на $\sqrt{n}$.

6. Для кратної вибірки з гамма-розподілом $\Gamma(1,\theta)$ спостережень скласти рівняння максимальної вірогідності та довести, що ОМВ є асимптотично нормальною та асимптотично ефективною. Знайти оцінку методу моментів та довести, що її ефективність менша за 1.

7. Кратна вибірка утворена з урізаного на інтервал (a,b) нормального розподілу $N(\theta,1)$:

$$f(y,\theta) = C(\theta,a,b)\exp(-(y-\theta)^2/2)\mathbb{1}_{y\in(a,b)}.$$

Довести, що вибіркове середнє $\widehat{\mu}_n$ є ОМВ для функції $\mu(\theta) = \mathbf{E}_\theta \xi_1$ та обчислити ОМВ для θ.

8. Знайти ОМВ параметра $\theta = (\theta_1, \theta_2)$ для кратної вибірки з розподілом спостережень

$$f(y,\theta) = \begin{cases} (\theta_1+\theta_2)^{-1}\exp\{-y/\theta_1\}, & y \ge 0, \\ (\theta_1+\theta_2)^{-1}\exp\{\;\; y/\theta_2\}, & y < 0, \end{cases} \qquad \theta > 0.$$

та обчислити її граничний розподіл.

9. Знайти оцінку максимальної вірогідності для кількості n можливих результатів у одному випробуванні в поліноміальній схемі з N випробувань при рівноймовірних результатах одного випробування за кількістю різних результатів.

10. Знайти оцінки максимальної вірогідності параметрів через кратну вибірку з гіпергеометричного розподілу.

11. Записати рівняння максимальної вірогідності параметрів для кратної вибірки з логістичною функцією розподілу спостережень $F(x) = (1 + \exp(-\alpha - \beta x))^{-1},\ x \in \mathbb{R}$.

12. Спостерігається кратна вибірка з двовимірними спостереженнями (ξ_{1j}, ξ_{2j}), причому (ξ_{11}, ξ_{21}) незалежні та мають показникові розподіли з параметрами λ_1, λ_2 відповідно. Знайти ОМВ для них. Припустимо, що замість вибірки спостерігаються величини $\eta_j = \min(\xi_{1j}, \xi_{2j})$ та $\delta_j = \mathbb{1}_{\eta_j = \xi_{1j}}$. Знайти ОМВ у цьому випадку.

13. Знайти ОМВ параметрів для кратної вибірки з від'ємно-біноміального розподілу: $\mathbf{P}(\xi_1 = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r},\ k \geq r$.

14. Знайти необхідну і достатню умову існування єдиного розв'язку рівняння максимальної вірогідності для кратної вибірки зі щільністю спостереження $f(y, \theta) = \theta f_0(y) + (1-\theta) f_1(y), y \in \mathbb{R}$. Вказати ОМВ при порушенні цієї умови.

15. Знайти асимптотичний розподіл оцінки максимальної вірогідності функції $\tau(\theta) = \theta(1-\theta)$ від імовірності успіху θ у схемі випробувань Бернуллі.

1.2.6. Інтервальні оцінки параметрів нормальних спостережень

Література: [1, с. 377–381], [2, с. 383–388]

А

1. Середній IQ (рівень інтелекту) 100 студентів університету Меріленду дорівнює 110 зі стандартним відхиленням 5. Знайти 95%-й довірчий інтервал для теоретичного середнього IQ генеральної сукупності всіх студентів університету.

2. Для визначення величини заряду електрона e_0 Міллікен зробив 58 незалежних однаково точних спостережень і в результаті отримав такі значення вибіркових середнього та дисперсії: $\widehat{\mu} = 4,7808$ та $\widehat{s}^2 = 23384 \cdot 10^{-8}$. Припускаючи, що 58 результатів спостережень – однаково нормально розподілені випадкові величини, обчислити довірчий інтервал для e_0 з коефіцієнтом надійності $0,975$.

3. Дві лабораторії визначали концентрацію сірки в дизельному паливі за стандартним зразком. Вісім вимірювань у першій лабораторії дали результати: 0,869; 0,874; 0,867; 0,875; 0,870; 0,869; 0,864; 0,872. Після десяти вимірювань другої лабораторії отримано: 0,865; 0,870; 0,866; 0,871; 0,868; 0,870; 0,871; 0,870; 0,869; 0,874. Істинне значення концентрації сірки в стандартному зразку дорівнює 0,870. Вважаючи систематичні похибки нульовими, знайти довірчі інтервали відношення двох невідомих дисперсій. Випадкові похибки вважаємо однаково розподіленими нормальними величинами. Коефіцієнт довіри становить $0,95$.

4. Випадкові величини $\bar{x}_1$, s_1^2 та $\bar{x}_2$, s_2^2 являють собою незміщені оцінки для математичних сподівань та дисперсій двох нормальних розподілів $N(a_1, \sigma_1^2)$ та $N(a_2, \sigma_2^2)$ відповідно, обчислені за вибірками об'ємами n_1 та n_2. У результаті експерименту виявилось, що при $n_1 = 21$ та $n_2 = 36$ ці оцінки набули значень: $\bar{x}_1 = 10.73$, $s_1^2 = 0.0221$ та $\bar{x}_2 = 10.17$, $s_2^2 = 0.382$. Знайти довірчий інтервал для різниці $a_1 - a_2$ за умови, що невідомі дисперсії нормальних розподілів однакові, тобто $\sigma_1^2 = \sigma_2^2$. Коефіцієнт довіри вибрати рівним 0.975.

5. Випадкова величина ξ_λ має розподіл Пуассона з параметром λ. Побудувати асимптотичний надійний інтервал для λ за спостереженням ξ_λ, виходячи з асимптотичної нормальності величини $(\xi_\lambda - \lambda)\,\lambda^{-1/2}$ при $\lambda \to \infty$.

B

1. Випадкова величина ξ має нормальний розподіл із відомим середнім квадратичним відхиленням $\sigma = 2$. Знайти 95%-й надійний інтервал для невідомого математичного сподівання ξ за вибірковим середнім $\widehat{\mu}$, якщо об'єм вибірки $n = 36$.

2. Із надійністю 95 % знайти довірчий інтервал для середнього квадратичного відхилення висоти внутрішніх кілець підшипників за даними вибірки об'єму $n = 25$, якщо їх розподіл по висоті нормальний і "виправлена" дисперсія $\widehat{s}^2 = 2.5282$.

3. П'ять незалежних вимірів однієї й тієї самої невідомої фізичної сталої μ дали результати 1.78, 1.81, 1.94, 1.86, 2.00. Для виявлення систематичної похибки аналогічним чином отримано п'ять результатів вимірювань 0.92, 0.89, 0.78, 0.82, 0.92 еталонного (відомого) значення $\mu_0 = 1$. Припускається, що систематична похибка не залежить від вимірюваного значення μ та що всі випадкові похибки взаємно незалежні і мають однаковий нормальний розподіл $N(0, \sigma^2)$ із невідомим додатним значенням параметра σ^2. Знайти для σ^2 надійні інтервали. Коефіцієнт надійності вибрати 0.95 для кожного.

4. Отримано дві випадкові вибірки з двох незалежних нормальних сукупностей (із середніми μ_1 і μ_2 відповідно) з такими результатами. Вибірка 1: об'єм вибірки $n_1 = 11$, вибіркові середнє $\overline{\mu}_1 = 124$ і дисперсія $s_1^2 = 59$. Вибірка 2: об'єм вибірки $n_2 = 15$, вибіркові середнє $\overline{\mu}_2 = 105$ і дисперсія $s_2^2 = 42$. Обчислити 95%-й вірогідний інтервал для різниці між математичними сподіваннями $\mu_1 - \mu_2$ (можна прийняти, що дисперсії сукупностей рівні).

5. Припустимо, що для статистики $T = T(X)$ імовірність $\mathbf{P}_\theta(T < x)$ монотонно зростає за θ при кожному x. Нехай для заданого рівня α функції θ_i визначено з рівнянь $\mathbf{P}_{\theta_1(x)}(T < x) = \frac{\alpha}{2}$ та $\mathbf{P}_{\theta_2(x)}(T \geq x) = \frac{\alpha}{2}$: а) довести, що $(\theta_1(T(X)), \theta_2(T(X)))$ є надійним інтервалом рівня $1 - \alpha$ для θ; б) аналогічні твердження вивести для монотонно спадної залежності. Знайти відповідні статистики $T(X)$ та перевірити, що вказане припущення виконується для біноміального та пуассонівського розподілів.

C

1. Вибірка складається з 2500 вимірювань температури повітря із середнім $\mu = 25$ °C і середнім квадратичним відхиленням $\sigma = 7$ °C. Яка ймовірність того, що середнє арифметичне вибірки об'ємом 100 лежить у межах: а) між 26 °C та 27 °C? б) між 22 °C та 24 °C?

2. Номінальна маса виробу становить 1 кг. Відомо, що 5 % виробів мають масу меншу, ніж 0,94 кг. Знайти ймовірність того, що з чотирьох виробів рівно два мають масу більшу, ніж 1,05 кг.

3. У результаті п'яти випробувань нормально розподіленої випадкової величини отримано такі значення: 2.96, 3.02, 3.07, 2.98, 3.06. Побудувати довірчі інтервали з коефіцієнтом надійності 0.95 кожен: а) для математичного сподівання нормального розподілу, якщо відома дисперсія $\sigma^2 = 0.01$; б) для дисперсії нормального розподілу, якщо відоме математичне сподівання $\mu = 3$; в) для математичного сподівання та дисперсії, якщо обидва невідомі.

4. Десять незалежних вимірювань невідомої величини дали такі результати: 2.96, 3.07, 3.02, 2.98, 3.06, 2.92, 2.88, 3.10, 3.06, 2.95. Припускаючи, що помилки вимірювань – однаково нормально розподілені випадкові величини і що систематичні помилки відсутні, указати для середнього та для невідомої дисперсії випадкових помилок довірчі інтервали з коефіцієнтом довіри 0.9.

5. Вивести з центральної граничної теореми, таке:

$$\left(\chi_n^2 - n\right)/\sqrt{2n} \xrightarrow{W} \zeta \simeq N(0,1),\ n \to \infty.$$

Вивести звідси вірогідний інтервал для невідомої кількості ступенів вільності n через одне спостереження $\xi = \chi_n^2$.

6. Довести апроксимацію Фішера:

$$\sqrt{2\chi_n^2} - \sqrt{2n-1} \xrightarrow{W} \zeta \simeq N(0,1),\ n \to \infty.$$

Вивести звідси вірогідний інтервал для невідомої кількості ступенів вільності n через одне спостереження $\xi = \chi_n^2$.

7. Довести, що для кратних нормальних спостережень вибіркові асиметрія та ексцес $\widehat{\mu}_{3n}^0/\widehat{s}_n^{3/2},\ \widehat{\mu}_{4n}^0/\widehat{s}_n^2$ не залежать від вибіркових середнього та дисперсії.

8. Виходячи з виразу для щільності хі-квадрат як окремого випадку гамма-щільності, довести, що щільність τ_n має вигляд

$f_{\tau_n}(x) = c_n(1+x^2/n)^{-(n+1)/2},\ c_n = (\pi n)^{-1/2}\Gamma((n+1)/2)/\Gamma(n/2)$, причому $\mathbf{E}\tau_n = 0,\ \mathbf{D}\tau_n = n/(n-2), n > 2$. Зокрема, τ_1 має розподіл Коші.

9. Довести, що величини з розподілом Стьюдента є асимптотично нормальними: $\tau_n \xrightarrow{W} \zeta \simeq N(0,1),\ n \to \infty$.

10. Довести, що щільність Фішера $\phi_{n,m}$ має вигляд

$f_{nm}(x) = c_{nm} x^{n/2-1} \left(1 + \frac{n}{m} x\right)^{-(n+m)/2},\ c_{nm} = (\frac{n}{m})^{n/2} B(\frac{n}{2}, \frac{m}{2})$.

11. Нехай f – строго монотонна функція така, що для значення $f(\theta)$ побудовано вірогідний інтервал $(\widehat{f_1}, \widehat{f_2})$. Довести, що $\left(f^{(-1)}(\widehat{f_1}), f^{(-1)}(\widehat{f_2})\right)$ є вірогідним інтервалом того самого рівня для θ.

12. Побудувати надійний інтервал заданого рівня для параметра кратної вибірки: а) з показниковим розподілом $Exp(\theta)$, б) рівномірним на $(\theta, 2\theta)$ розподілом, в) розподілом Пуассона.

13. 300 випробувань Бернуллі з невідомою ймовірністю успіху θ призвели до 164 успіхів. Побудувати вірогідний інтервал для дисперсії кількості успіхів при надійності 0.95.

14. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна нормальна вибірка з розподілом $\xi_k \simeq N(\mu, \sigma^2)$. Знайти вірогідний інтервал рівня p для наступного незалежного спостереження ξ_{n+1}.

15. Нехай $X = (\xi_1, ..., \xi_n)$ – кратна нормальна вибірка з розподілом $\xi_k \simeq N(\mu, \sigma^2)$, $\Phi_{\mu\sigma}$ – нормальна функція розподілу з параметрами μ, σ^2. Довести: а) для кожного $t > 0$ розподіл приросту

$$\Delta_t = \Phi_{\mu\sigma}(\widehat{\mu}_n + t\widehat{\sigma}_n) - \Phi_{\mu\sigma}(\widehat{\mu}_n - t\widehat{\sigma}_n)$$

залежить лише від t; б) якщо $t = y_{n-1,\alpha}\sqrt{1 + 1/n}$, то $\mathbf{E}\Delta_t = \alpha$; в) $\sqrt{n}(\widehat{\mu}_n/\widehat{s}_n - \kappa)/\sqrt{1 + \kappa^2/2} \xrightarrow{W} N(0,1), n \to \infty$, де $\kappa = \mu/\sigma$.

16. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка з нормальним розподілом на площині: $\xi_1 \simeq N_2(0, 0, 1, 1, \rho)$. Знайти надійний інтервал для ρ при заданому вірогідному рівні α.

17. Кратні вибірки $X = (\xi_1, \ldots, \xi_n)$ та $Y = (\eta_1, \ldots, \eta_m)$ містять спостереження з показниковими розподілами $Exp(\lambda)$ та $Exp(\mu)$ відповідно. Знайти надійний інтервал рівня p для відношення λ/μ.

18. Побудувати вірогідний інтервал рівня p для параметра λ за кратною вибіркою з гамма-розподілу $\Gamma(\lambda, \alpha)$ при відомому α.

19. Довести, що інтервал $\left(\xi_{(n)}, \xi_{(n)}(1-p)^{-1/n}\right)$ для параметра θ за кратною вибіркою з рівномірного розподілу $U(0, \theta)$ має вірогідний рівень p.

1.3. Перевірка статистичних гіпотез

1.3.1. Непараметричні критерії узгодженості, однорідності, незалежності

Література: [1, с. 386–394], [2, с. 389–401]

А

1. Наведені нижче дані являють собою час безвідмовної роботи електронного приладу. З'ясувати, чи можна вважати, що час безвідмовної роботи цього приладу має нормальний розподіл із математичним сподіванням 170 годин та середнім квадратичним відхиленням, яке дорівнює 110 годин? Час безвідмовної роботи: 48, 84, 21, 63, 219, 133, 327, 213, 118, 171, 356, 117, 175, 163, 230, 135, 18, 116, 174, 384, 100, 60, 327, 48, 410 *(критерій Колмогорова)*.

2. У першому потоці з 300 студентів оцінку "2" отримали 33 студенти, "3" – 43, "4" – 80, "5" – 144, у другому потоці інші 300 студентів отримали такі результати: "2" – 39, "3" – 35, "4" – 72, "5" – 154. Чи можна з вірогідним рівнем 0.05 вважати обидва потоки однорідними? (Інакше кажучи, чи однаково розподілені дві величини, значення яких указані в цій задачі? *Критерій Вілкоксона.*)

3. Довести, що для статистики Вілкоксона у припущенні неперервності розподілу спостережень

$$S_{nm} = \frac{n(n+1)}{2} + \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{1}_{\{\xi_i > \eta_j\}}.$$

Обчислити звідси $\mathbf{E}S_{mn}$, $\mathbf{D}S_{mn}$ за альтернативною гіпотезою зсуву однієї вибірки відносно іншої.

4. Дані про смертність населення використовуються для оцінювання параметрів захворюваності. За наведеними нижче даними для 10 регіонів перевірити гіпотезу про (не)залежність смертності від захворюваності за критерієм Спірмена:

Регіон	Смертність на 10000	Захворюваність на 1000
1	125.2	206.8
2	119.3	213.8
3	125.3	197.2
4	111.7	200.6
5	117.3	189.1
6	100.7	183.6
7	108.8	181.2
8	102.0	168.2
9	104.7	165.2
10	121.1	228.5

B

1. Проби дуже чистого заліза, одержаного за двома різними методами A та B, мали такі точки плавлення:

A: 1493, 1519, 1518, 1512, 1514, 1489, 1508, 1503
B: 1509, 1494, 1512, 1483, 1507, 1491

Перевірити гіпотезу, згідно з якою обидва методи дають залізо, що має одну й ту саму точку плавлення (*критерій Вілкоксона*).

2. За даними задачі А.4 за допомогою критерію Смірнова перевірити гіпотезу про однаковий розподіл захворюваності та смертності за регіонами, використовуючи нормування значень захворюваності з коефіцієнтом 1,7.

3. Довести що в умовах теореми Колмогорова про відхилення емпіричної функції розподілу за нульової гіпотези розподіл статистики омега-квадрат $\widehat{\omega}_n^2 = n \int\limits_{-\infty}^{\infty} \left(\widehat{F}_n(x) - F(x)\right)^2 dF(x)$ не залежить від вигляду теоретичної функції розподілу.

4. На підставі виразів для коваріацій рангових статистик перевірити таку рівність для статистики Спірмена: $\mathbf{D}\widehat{\rho}_n = 1/(n-1)$ за основної гіпотези H_0.

C

1. Обчислити дисперсію статистики Вілкоксона.

2. Проведено 10 незалежних випробувань гіпотези з 5%-м критичним рівнем. Обчислити ймовірність того, що принаймні одне з випробувань приведе до істотного результату у припущенні, що нульова гіпотеза справджується для кожного з 10 випробувань.

3. Студентський актуарій збирається зібрати дані з партії форм страхування. Відомо, що в середньому одна з кожних п'ятдесяти таких форм містить неповну інформацію. Студент хоче бути на 95 % упевненим щодо наявності принаймні 500 форм, які є повними. Показати, що він повинен використати партію принаймні з 516 форм. (Використати придатну модель та відповідні таблиці для розподілу кількості форм, що містять неповну інформацію.)

4. Спостерігаються дві незалежні схеми випробувань Бернуллі з n_i спостереженнями та ймовірностями успіху $\theta_i, i = 1, 2$. Нехай $\widehat{\theta}_i$ – відповідні кількості успіхів. Довести, що при $n_i \to \infty$:

$$\frac{\left(\widehat{\theta}_1 - \widehat{\theta}_2 - \theta_1 + \theta_2\right)}{\sqrt{\theta_1(1-\theta_1)/n_1 + \theta_2(1-\theta_2)/n_2}} \xrightarrow{W} \zeta \simeq N(0,1),$$

а за умови $H_0\colon \theta_1 = \theta_2$ також

$$\frac{\left(\widehat{\theta}_1 - \widehat{\theta}_2\right)}{\sqrt{\widehat{\theta}(1-\widehat{\theta})(1/n_1 + 1/n_2)}} \xrightarrow{W} \zeta \simeq N(0,1),$$

де $\widehat{\theta} = \left(n_1\widehat{\theta}_1 + n_2\widehat{\theta}_2\right)/(n_1 + n_2)$. Використати ці твердження для перевірки гіпотези H_0.

5. Вивести з означення нормованої статистики Вілкоксона $\varkappa_{nm} = (S_{nm} - \mathbf{E}S_{nm})/\sqrt{\mathbf{D}S_{nm}}$, що за нульової гіпотези

$$\mathbf{E}(\varkappa_{nm})^{2r-1} = 0, \quad \mathbf{E}(\varkappa_{nm})^{2r} \to (2r-1)!!,\ n \to \infty,$$

при всіх $r \geq 1$, та довести асимптотичну нормальність $\varkappa_{nm}$.

6. Довести асимптотичну нормальність статистики Спірмена $\widehat{\rho}_n$ обчисленням моментів $\mathbf{E}\widehat{\rho}_n^{2r}$ при $r \geq 1$ та їх границь при $n \to \infty$.

7. Нехай $\widehat{F}_n(x)$ – емпірична функція розподілу для вибірки з неперервною функцією розподілу F одного спостереження, $\alpha > 0$, а функція $w(t)$ невід'ємна. Довести, що розподіл статистик

$$S_{nw\alpha} = \sup_{x\in\mathbb{R}} w(F(x)) \left|\widehat{F}_n(x) - F(x)\right|^{\alpha},$$
$$T_{nw\alpha} = \sup_{x\in\mathbb{R}} w\left(\widehat{F}_n(x)\right) \left|\widehat{F}_n(x) - F(x)\right|^{\alpha},$$
$$U_{nw\alpha} = \int_{\mathbb{R}} w(F(x)) \left|\widehat{F}_n(x) - F(x)\right|^{\alpha} F(dx),$$
$$V_{nw\alpha} = \int_{\mathbb{R}} w\left(\widehat{F}(x)\right) \left|\widehat{F}_n(x) - F(x)\right|^{\alpha} \widehat{F}(dx)$$

не залежить від F.

8. Обчислити асимптотичну ефективність критерію Вілкоксона відносно t-критерію Стьюдента у випадку кратної вибірки: а) з логістичного розподілу $F(x) = (1 + \exp(-\alpha - \beta x))^{-1}$, б) з рівномірного розподілу $U(a, b)$ при альтернативі зсуву.

1.3.2. Критерії хі-квадрат для поліноміальної схеми Бернуллі

Література: [1, с. 395–405], [2, с. 402–413]

А

1. Чотири монети підкинуто 20160 разів, при цьому комбінації: чотири аверси, три аверси і реверс, два аверси і два реверси, один аверс і три реверси, чотири реверси – з'явилися відповідно таку кількість разів: 1181, 4909, 7583, 5085, 1402. Чи свідчать ці дані про те, що кількість аверсів, що з'явилися в результаті підкидання чотирьох монет, є біноміально розподіленою випадковою величиною з параметрами 4, 1/2?

2. Знайти сумісний розподіл та його генератрису для вектора емпіричних частот $\widehat{\nu}_n = (\widehat{\nu}_{n1}, ..., \widehat{\nu}_{nk})$ у поліноміальній схемі Бернуллі.

3. Нижче наведено інтервали в експлуатаційних годинах між послідовними відмовами апаратури кондиціювання повітря на літаку "Боїнг-720": 6, 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 4, 11, 16, 90, 1, 16, 52, 95, 97, 51, 11, 4, 141, 18, 142, 68, 77, 80, 1, 16, 106, 206, 82, 54, 31, 216, 46, 111, 39, 63, 18, 191, 18, 163, 24, 50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14, 487, 18, 100, 7, 98, 5, 85, 91, 43, 230, 3, 130. Перевірте гіпотезу про показниковий

розподіл із параметром 0.01 часу безвідмовної роботи апаратури кондиціювання повітря.

4. У результаті щосекундної реєстрації кількості альфа-частинок, що потрапили до камери лічильника, отримано такі дані:

Кількість частинок	0	1	2	3	4	5	6	≥ 7
Кількість секунд	112	168	130	68	32	5	2	0

Перевірити гіпотезу про пуассонівський розподіл кількості частинок із параметром 1.5 на рівні 0.05.

B

1. Результати підрахунку частот цифр $0, 1, \ldots, 9$ у перших 10002 десяткових знаках числа $\pi - 3$ такі:

0	1	2	3	4	5	6	7	8	9
968	1026	1021	974	1014	1046	1021	970	948	1014

За допомогою критерію хі-квадрат на рівні 0.01 перевірити гіпотезу про рівномірну розподіленість десяткових цифр числа π.

2. Нехай $\widehat{\nu}_{nij}$ – кількість спостережень j-го результату до першої появи i-го результату у поліноміальній схемі Бернуллі. Знайти: а) генератрису, математичне сподівання та дисперсію величини $\widehat{\nu}_{nij}$; б) знайти сумісну генератрису вектора $(\widehat{\nu}_{nij}, j = \overline{1, i-1})$.

3. Оберіть одну з таблиць наприкінці цього посібника, із перших 100 її чисел випишіть другу справа цифру та перевірте гіпотезу про випадковість цих цифр. Чому не бажано обирати останню цифру ?

4. Із 1000 застрахованих автовласників за один рік 123 потрапили у ДТП по одному разу, 22 – по два рази. При вірогідному рівні 0,01 перевірити гіпотезу про те, що кількість ДТП має розподіл Пуассона з параметром 0,1.

C

1. Довести, що статистика хі-квадрат у поліноміальній схемі Бернуллі така: $\widehat{\chi}^2_{k-1}(n) = \sum\limits_{i=1}^{k} \widehat{\nu}^2_{ni}/np_i - n$.

2. В експериментах із селекцією гороху Мендель спостерігав частоти появи різних видів насіння, отриманих у результаті схрещування рослин із круглим жовтим насінням і рослин зі зморшкуватим зеленим насінням:

Таблиця 1.3.1

Насіння	**Частота**	**Імовірність**
кругле жовте	315	$9/16$
зморшкувате жовте	101	$3/16$
кругле зелене	108	$3/16$
зморшкувате зелене	32	$1/16$
Всього	556	1

Перевірити гіпотезу про відповідність спостережень теоретичним частотам.

3. У серії з 4000 незалежних випробувань події A_1, A_2, A_3 з повної групи подій відбулись відповідно 1905, 1015 та 1080 раз. Перевірити на рівні 0.05 гіпотезу про те, що ймовірності цих подій відповідно дорівнюють $1/2, 1/4, 1/4$.

4. Серед 1110 сімей, що мають двох дітей, у 527 сім'ях два хлопчики, у 476 – дві дівчинки, а у решти 107 сімей по одному хлопчику та дівчинці. Чи можна на рівні 0.05 вважати, що кількість хлопчиків у сім'ї з двома дітьми має біноміальний розподіл із параметрами 2 та $1/2$?

5. Обстеження 500 годинників, що виставлені у вітринах у центрі міста, виявило такі результати щодо положення годинникової стрілки:

Година	0	1	2	3	4	5	6	7	8	9	10	11
Частота	41	34	54	39	49	45	41	33	37	41	47	39

Перевірити гіпотезу про рівномірний розподіл початкової години в обстежених годинників.

6. Довести, що заміна повної групи подій у поліноміальній схемі випробувань шляхом об'єднання деяких гіпотез не приводить до збільшення статистики хі-квадрат.

1.3.3. Критерії хі-квадрат для складної гіпотези

Література: [1, с. 395–405], [2, с. 402–413]

А

1. Серед 1110 сімей, що мають двох дітей, у 527 сім'ях два хлопчики, у 476 – дві дівчинки, а у решти 107 сімей по одному хлопчику та дівчинці. Чи можна на рівні 0.05 вважати, що кількість хлопчиків у сім'ї з двома дітьми має біноміальний розподіл?

2. 147 навмання вибраних студентів були поділені згідно з кольором їх волосся (біляві, темні) та кольором їх очей (блакитні, карі).

Таблиця 1.3.2

Колір волосся	**Колір очей**		
	Блакитні	Карі	Всього
Темні	31	41	72
Біляві	40	35	75
Всього	71	76	147

Чи можна на підставі цих даних зробити висновок про те, що колір очей пов'язаний із кольором волосся?

3. Із випадково обраних 250 осіб, що переглядали один телевізійний канал, 74 пригадали зміст реклами, а з 250 осіб, що переглядали інший телеканал, зміст реклами пригадали 92 особи. Перевірити гіпотезу про однакову дієвість реклами на каналах на вірогідному рівні 0.01.

4. Побудувати критерій хі-квадрат узгодженості рівня p кратної вибірки $X = (\xi_1, \ldots, \xi_n)$ із показниковою функцією розподілу $F_\xi(x) = 1 - \exp(-x/\theta)$ при невідомому параметрі θ для N інтервалів групування довжини a (за винятком останнього).

В

1. Нижче наведено інтервали в експлуатаційних годинах між послідовними відмовами апаратури кондиціювання повітря на літаку “Боїнг-720”: 6, 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42,

20, 5, 12, 120, 11, 3, 14, 71, 11, 4, 11, 16, 90, 1, 16, 52, 95, 97, 51, 11, 4, 141, 18, 142, 68, 77, 80, 1, 16, 106, 206, 82, 54, 31, 216, 46, 111, 39, 63, 18, 191, 18, 163, 24, 50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14, 487, 18, 100, 7, 98, 5, 85, 91, 43, 230, 3, 130. Перевірити гіпотезу про показниковий розподіл часу безвідмовної роботи апаратури кондиціювання повітря.

2. Психолог припускає, що може існувати зв'язок між характером та улюбленим кольором людини. У результаті опитування 150 інтровертів та 150 екстравертів він отримав такі результати:

Таблиця 1.3.3

Психологічний тип	**Червоний**	**Синій**	**Фіолетовий**	**Жовтий**	**Зелений**	**Всього**
Інтроверт	52	28	19	38	13	150
Екстраверт	48	32	21	32	17	150
Всього	100	60	40	70	30	300

Перевірити цю гіпотезу на вірогідному рівні 0.05.

3. Довести, що у (2×2)-таблиці спряженості умовний розподіл частоти $\widehat{\nu}_{11}$ за умови, що фіксовані $\widehat{\nu}_{\bullet j}$ та $\widehat{\nu}_{i\bullet}$, є гіпергеометричним. Знайти його. Побудувати точний критерій Фішера перевірки незалежності, якщо критичними є великі відхилення в обидва боки величини $\widehat{\nu}_{11}$ від $\widehat{\nu}_{\bullet 1}\widehat{\nu}_{1\bullet}/n$.

4. У генетичній моделі Фішера припускається, що ймовірності появи нащадків, що класифікуються за чотирма типами, мають вигляд: $p_1(\theta) = (2+\theta)/4$, $p_2(\theta) = p_3(\theta) = (1-\theta)/4$, $p_4(\theta) = \theta/4$, де $\theta \in (0,1)$ – невідомий параметр. Побудувати критерій хі-квадрат рівня p для перевірки цієї гіпотези на підставі кратної вибірки.

C

1. Розглядається (2×2)-таблиця (табл. 1.3.4) спряженості для перевірки незалежності двох факторів:

Таблиця 1.3.4

Фактор	**Частота**		
1	22	28	50
2	28	22	50
Всього	50	50	100

Обчислити значення статистики хі-квадрат та сформулювати висновок щодо незалежності факторів.

2. Перевірити гіпотезу про пуассонівський розподіл у задачі А4 з розділу 1.3.2 при невідомому параметрі θ.

3. Із 300 абітурієнтів 97 мали вищий бал шкільного атестата і 48 отримали вищий бал на вступних іспитах. Лише у 18 з них був вищий бал атестата і на іспитах одночасно. Побудувати таблицю спряженості і перевірити на рівні 0.1 гіпотезу про незалежність вищих балів.

4. Під час епідемії грипу з 3000 осіб захворіли на грип 273 особи, із них 14 хворіли двічі. Решта осіб не захворіли. Перевірити гіпотезу про те, що розподіл кратності захворювання на грип є пуассонівським.

5. У 8000 незалежних випробуваннях події A, B, C, що утворюють повну групу, спостерігались відповідно 2014, 5012, та 974 рази. Перевірити на рівні 0.05 гіпотезу про те, що $\mathbf{P}(A) = 0.5 - 2\theta$, $\mathbf{P}(B) = 0.5 + \theta$, $\mathbf{P}(C) = \theta$, де $\theta \in (0, 0.25)$ – невідомий параметр.

6. Перевірити гіпотезу про нормальний розподіл часу безвідмовної роботи у задачі В.1.

7. Кількість вимог у портфелі страхових полісів відображено в такий спосіб:

Кількість вимог за день	0	1	2	3	4	5	Всього
Частота вимог	48	32	17	2	0	1	100

Використати критерій хі-квадрат для перевірки гіпотези про те, що кількість вимог має розподіл Пуассона.

8. Дослідник, що вивчає досвід вимог компанії (у конкретному класі бізнесу), реєструє платежі за 100 вимогами. Платежі (у тисячах, сортовані) наведено нижче:

0.30	0.89	0.96	1.16	1.67	1.77
1.93	1.98	2.07	2.09	2.30	2.48
2.58	2.78	3.00	3.19	3.21	3.21
3.25	3.31	3.34	3.37	3.66	3.95
4.16	4.18	4.60	4.72	4.73	4.76
5.01	5.17	5.21	5.63	5.72	6.00
6.13	6.17	6.24	6.37	6.47	6.48
6.87	7.05	7.16	7.21	7.51	7.72
7.74	8.00	8.00	8.03	8.04	8.54
9.11	9.18	9.49	9.59	10.00	10.36
10.85	11.08	11.22	11.27	11.38	11.45
11.69	11.78	12.27	12.30	12.50	13.04
13.28	13.43	13.48	13.85	14.27	14.31
14.49	14.55	14.62	14.68	14.70	14.83
15.67	15.70	15.77	16.28	16.44	17.17
17.89	18.03	18.12	20.72	22.00	24.33
25.41	28.30	31.00	32.80		

Для цих 100 спостережень: $\sum x = 952.75$, $\sum x^2 = 13584.5217$.

Дослідник хоче перевірити, чи показниковий розподіл забезпечує адекватний опис розподілу платежів: а) обчислити вибіркове середнє і стандартне відхилення для 100 платежів; б) визначити оцінку параметра показникового розподілу; в) дослідник вирішив провести тест хі-квадрат для показникового розподілу даних, використовуючи п'ять рівноймовірних інтервалів (тобто інтервалів з відповідною ймовірністю 0.2). Показати, що значення x, яке перевищується з імовірністю p показниковою величиною із середнім μ, таке: $x = -\mu \ln p$; г) обчислити значення, які ділять множину додатних дійсних чисел на п'ять рівноймовірних інтервалів для оціненого показникового розподілу.

9. Перевірка 190 видів радіоприймачів виявила такі параметри чутливості та вибірковості (табл. 1.3.5):

Таблиця 1.3.5

Чутливість	**Вибірковість**		
	Низька	Середня	Висока
Низька	7	12	31
Середня	35	59	18
Висока	15	13	0

На вірогідному рівні 0.01 перевірити гіпотезу про незалежність чутливості від вибірковості.

10. Із 1000 застрахованих автовласників за один рік 123 потрапили у ДТП по одному разу, 22 – двічі. При вірогідному рівні 0.99 перевірити гіпотезу про те, що кількість ДТП має розподіл Пуассона.

11. Соціальний дослідник цікавиться розподілом статі серед дітей у сім'ях, де є діти, і зібрав дані для дослідження у такий спосіб. 300 сімей були відібрані навмання. Частоти розподілу дівчаток у сім'ях, що мають 1, 2, 3 і 4 дитини, наведено в табл. 1.3.6.

Таблиця 1.3.6

Кількість дітей	**Кількість дівчаток**					**Кількість сімей**
	0	1	2	3	4	
1	23	27	-	-	-	50
2	30	46	24	-	-	100
3	9	36	43	12	-	100
4	4	17	15	11	3	50

Дослідник хоче з'ясувати, чи залежить пропорція кількості дівчаток у межах сім'ї від розміру сім'ї: а) побудувати відповідну (2×4)-таблицю спряженості та обчислити повну пропорцію дівчаток; б) сформулювати відповідні гіпотези для подальшого дослідження; в) обчислити значення відповідних тестів та перевірити, чи їх P-значення перевищують 0.05.

12. Статистика кількості страхових випадків протягом часу дії 576 страхових полісів дає:

Кількість випадків	0	1	2	3	4	5	≥ 6
Кількість полісів	229	211	93	35	7	1	0

Перевірити гіпотезу про пуассонівський розподіл кількості страхових випадків на рівні 0.05.

13. Довести, що статистика для перевірки гіпотези про симетрію $H_0\colon\ p_{ij} = p_{ji},\ i, j = \overline{1, r}$, записана у вигляді

$$\chi^2(r) = \sum_{i<j} (\upsilon_{ij} - \upsilon_{ji})^2 / (\upsilon_{ij} + \upsilon_{ji}),$$

має асимптотичний розподіл $\chi^2_{r(r-1)/2}$.

1.3.4. Перевірка гіпотез про параметри нормальних спостережень

Література: [1, с. 406–411], [2, с. 413–419]

А

1. Номінальний опір резисторів 2000 Ом. Для контролю добрана партія з 12 резисторів. Після вимірювання опору кожного зразка із середньоквадратичним відхиленням 5 Ом отримано такі значення: 2130, 2090, 2030, 2080, 1920, 2020, 2015, 2000, 2045, 1940, 1980, 1970. Чи можна відхилення від номіналу (2000 Ом) розглядати як випадкові (припустимі), чи, навпаки, результати свідчать про те, що опір резисторів істотно відрізняється від номіналу? Використати нормальну модель.

2. У результаті вимірювання 10 зразків твердості сплаву отримано такі значення в умовних одиницях: 12.1, 13.7, 11.0, 11.6, 11.9, 13.9, 11.5, 12.9, 13.0, 10.5. Припустимо, що твердість сплаву розподілена нормально. Чи можна вважати, що дисперсія розподілу твердості становить 2.25?

3. Дві лабораторії визначали концентрацію сірки в дизельному паливі за стандартним зразком. Вісім вимірювань у першій лабораторії дали результати:

0.869, 0.874, 0.867, 0.875, 0.870, 0.869, 0.864, 0.872.

Після десяти вимірювань у другій лабораторії отримано:

0.865, 0.870, 0.866, 0.871, 0.868, 0.870, 0.871, 0.870, 0.869, 0.874.

Істинне значення концентрації сірки в стандартному зразку дорівнює 0.870. Вважаючи, що випадкові похибки вимірювань нормально розподілені з однаковими дисперсіями, перевірити гіпотезу про те, що похибки в обох лабораторіях однакові. Перевірити гіпотезу, про те, що дисперсії похибок вимірювань однакові.

4. Нормальні випадкові величини ξ_k обчислюються із системи

$$\xi_k = \theta\xi_{k-1} + \varepsilon_k,\ k = \overline{1,n},$$

де похибки незалежні та $\varepsilon_k \simeq N(0,\sigma^2)$ і $\xi_0 = 0$. Довести, що сумісна щільність вектора $X = (\xi_1, \ldots, \xi_n)$ така:

$$L(x,\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{k=1}^{n}(x_k - \theta x_{k-1})^2/2\sigma^2\right),$$

де $x_0 = 0$. Довести, що відношення вірогідностей $L(X,\theta_1)/L(X,\theta_2)$ визначається статистиками $\sum\limits_{k=1}^{n} \xi_k\xi_{k-1}$ та $\sum\limits_{k=1}^{n} \xi_{k-1}^2$.

B

1. 12 хлопчиків та 10 дівчаток пройшли тест на швидкість читання, у результаті якого виявилось, що середня швидкість читання у хлопчиків дорівнює 74 слів за хвилину зі стандартним відхиленням 3, а у дівчаток – 79 зі стандартним відхиленням 4. Чи можна вважати що здібності до читання у хлопчиків та дівчаток відрізняються з вірогідним рівнем 0.05?

2. При порівнянні середніх страхових премій для полісів двох компаній використано двовибірковий t-тест у припущенні, що дисперсії вибірок однакові. Об'єми вибірок та вибіркові дисперсії: $n_1 = 25$, $\sigma_1^2 = 139.7$, $n_2 = 30$, $\sigma_2^2 = 76.6$. Провести відповідний F-тест на 5%-му рівні для перевірки гіпотези про рівність дисперсій вибірок.

3. Вибіркове середнє для 16 спостережень певної випадкової величини дорівнює 2.4, а вибіркове стандартне відхилення для даної вибірки дорівнює 0.2. Чи можна стверджувати, що вибірка одержана з нормального розподілу із середнім 2.5 (вірогідний рівень $\alpha = 0.05$)?

4. Розглядаються дві незалежні кратні вибірки об'єму n із нормальними розподілами спостережень, що мають невідомі середні μ_k та відомі дисперсії σ_k^2, $k = 1, 2$. Довести, що мінімальна кількість спостережень для перевірки гіпотези $H_0\colon \mu_1 - \mu_2 = 0$ проти альтернативи $H_1 : \mu_1 - \mu_2 = \delta$ при похибках α, β першого та другого роду, є $n_{\alpha\beta} = (\sigma_1^2 + \sigma_2^2)(x_\alpha + x_\beta)^2/\delta^2$, де x_α, x_β – квантилі рівнів $1 - \alpha, 1 - \beta$ для стандартного нормального розподілу.

C

1. Зі статистичних опитувань 800 домовласників у США з'ясувалось, що 29 % із них мають ступінь бакалавра. Оцінити на вірогідному рівні 0.05 теоретичну ймовірність того, що серед 4 наступних опитуваних є хоча б один бакалавр.

2. У результаті опитування 17 працівників компанії A та 25 працівників компанії B виявилось, що середня заробітна плата працівників компанії A становить 10000 зі стандартним відхиленням 900, а середня заробітна плата працівників компанії B становить 9000 зі стандартним відхиленням 800. Вважаючи розмір заробітної плати нормально розподіленою випадковою величиною, перевірити з вірогідним рівнем 0.01 гіпотезу про те, що дисперсія зарплати працівників компанії A більша, ніж дисперсія зарплати в компанії B.

3. Унаслідок вимірювання зросту 12 дорослих японців та 12 дорослих англійців визначили, що середній зріст японців становить 167.5 см із середнім квадратичним відхиленням 7.5 см, а середній зріст англійців становить 175 см із середнім квадратичним відхиленням 5 см. Чи можна стверджувати з вірогідним рівнем 0.05, що дорослі англійці вищі, ніж японці?

4. Спостерігається кратна вибірка $X = (\xi_k, k = \overline{1,n})$ із нормального розподілу $N(\mu, \sigma^2)$, де параметри μ та σ^2 – невідомі. Побудувати на підставі $t-$статистики $\sqrt{n}\widehat{\mu}_n/\widehat{s}_n$ критерій рівня α для перевірки гіпотези $\mu/\sigma \le \theta_0$ проти альтернативи $\mu/\sigma > \theta_0$ та дослідити його потужність.

5. Спостерігається кратна вибірка $X = ((\xi_{1k}, \xi_{2k}), k = \overline{1,n})$ із двовимірного нормального розподілу $N_2(0, 0, \sigma_1^2, \sigma_2^2, \rho)$. Довести, що статистика критерію відношення вірогідностей для перевірки

гіпотези $H_0\colon \rho = 0$ проти альтернативи $H_1\colon \rho \neq 0$ є функцією від $|\tau|$, де

$$\tau = \left(\sum_{k=1}^{n} \xi_{1k}\xi_{2k}\right) \Big/ \sqrt{\left(\sum_{k=1}^{n} \xi_{1k}^2\right)\left(\sum_{k=1}^{n} \xi_{2k}^2\right)}.$$

1.3.5. Найбільш потужні критерії відношення вірогідностей

Література: [1, с. 412–422], [2, с. 419–427]

А

1. Нехай про невідому ймовірність "успіху" у схемі Бернуллі $B(1,\theta)$ є дві гіпотези: $H_0\colon \theta = \theta_0$ та $H_1\colon \theta = \theta_1 > \theta_0$. Побудувати критерій Неймана —Пірсона для гіпотези H_0 проти альтернативи H_1. Знайти потужність цього критерію.

2. Кількість вимог, які виникають через рік для полісу деякого типу, моделюється розподілом Пуассона із середнім λ. Необхідно перевірити гіпотезу $H_0\colon \lambda = 0.2$ проти $H_1\colon \lambda > 0.2$. Вирішено відхилити H_0 на користь H_1, якщо 15 або більше вимог виникають через рік для групи з 50 незалежних таких полісів. Використовуючи асимптотичну нормальність розподілу Пуассона, обчислити потужність цього тесту в кожному з випадків: а) $\lambda = 0.3$ і б) $\lambda = 0.4$.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з функцією розподілу спостережень F. Побудувати критерій відношення вірогідностей для перевірки гіпотези про показниковий розподіл $F_0(x) = 1 - \exp(-x)$ проти альтернативи $F_1(x) = 1 - \exp(-\theta x), x \geq 0, \theta > 1$.

4. N осіб проходять аналіз крові, з імовірністю p фіксується захворювання кожної незалежно від інших. Процедура перевірки зводиться до аналізу сумарної проби кожної з k підгруп із приблизно однаковою кількістю осіб у кожній. Якщо результат негативний (жодна особа з підгрупи не захворіла), то він поширюється на кожну особу з підгрупи, в іншому випадку – повторно аналізується кожна особа підгрупи. Довести, що при малих p значення k, що мінімізує середню кількість тестів, наближено дорівнює

$N\sqrt{p}$, а відповідна кількість тестів становить $2N\sqrt{p}$. Порівняйте цю кількість із відповідною кількістю для стандартної процедури.

B

1. Нехай для розподілу Коші (θ) перевіряється гіпотеза $H_0\colon \theta = 0$ проти альтернативи $H_1\colon \theta = 1$. Показати, що рівномірно найпотужніший критерій вірогідного рівня $\alpha = \frac{1}{2} - \frac{1}{\pi}\operatorname{arctg}\left(\frac{1}{2}\right) \approx$ $\approx 0,352$ за одним спостереженням задається вибірковою критичною областю $\left\{x \geq \frac{1}{2}\right\}$ та що його потужність дорівнює $\frac{1}{2}+$ $+\frac{1}{\pi}\operatorname{arctg}\left(\frac{1}{2}\right) \approx 0,648$. Якщо вірогідний рівень $\alpha = \frac{1}{\pi}(\operatorname{arctg} 3-$ $-\operatorname{arctg} 1) \approx 0,148$, то вибіркова критична область має вигляд $\{1 \leq x \leq 3\}$, потужність критерію дорівнює $\frac{1}{\pi}\operatorname{arctg} 2 \approx 0,352$.

2. Знайти критерій відношення вірогідностей для нормальної схеми за кратною вибіркою для перевірки нульової та альтернативної гіпотез вигляду: $H_i\colon \mu = \mu_i, \sigma = \sigma_i, i = 0,1$, із відомими дисперсіями σ_i^2 (використати теорему про розподіл вибіркових статистик для нормальної вибірки).

3. Обчислити критерій відношення вірогідностей для перевірки гіпотези про значення параметра для кратної вибірки: а) з пуассонівським, б) геометричним, в) показниковим розподілом. Знайти потужність критерію, перевірити його незміщеність та консистентність.

4. Нехай $X = (\xi_1, \ldots, \xi_n)$ – кратна вибірка, а $f(y,\theta)$ – функція вірогідності одного спостереження:
а) довести, що $\ln l_{01}(X) = \sum\limits_{k=1}^{n} \eta_k$, де випадкові величини $\eta_k = \ln(f(\xi_k,\theta_1)/f(\xi_k,\theta_0))$ незалежні та однаково розподілені;
б) $\mathbf{E}_{\theta_0}\eta_1 = -I(\theta_1,\theta_0)$, де $I(\theta_1,\theta_0)$ – інформація за Кульбаком;
в) довести, що при $n \to \infty$ для критичного рівня в лемі Неймана – Пірсона виконується зображення $\ln l_\alpha + nI(\theta_1,\theta_0) \sim x_\alpha \sigma_0 \sqrt{n}$, де $\sigma_0^2 = \mathbf{D}_{\theta_0}\eta_1,\ \Phi(x_\alpha) = 1 - \alpha$;
г) вивести асимптотичне зображення для ймовірності похибки другого роду.

C

1. Довести, що оптимальні а) байєсівський та б) мінімаксний критерії для перевірки простих гіпотез також основані на статистиці відношення вірогідностей.

2. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка зі щільністю оберненого нормального розподілу:

$$f(y, \theta) = (\theta/2\pi y^3)^{1/2} \exp(-\theta y/2 + \theta - \theta/y), y > 0.$$

Знайти критерій Неймана – Пірсона для перевірки гіпотези $H_0\colon \theta = \theta_0$ проти альтернативи $H_1\colon \theta < \theta_0$, на основі нормальної апроксимації.

3. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з нормального розподілу $N(\theta, \sigma^2)$, де σ^2 – відома дисперсія. Побудувати критерій Неймана – Пірсона перевірки гіпотези $H_0\colon \theta = \theta_0$ проти альтернативи $H_1\colon \ \theta = \theta_1 > \theta_0$. Знайти потужність побудованого критерію.

4. Спостерігається вибірка $\xi_1, \ldots, \xi_n$ з незалежних спостережень, що мають розподіли Пуассона: $\xi_j \simeq \Pi(\lambda_j)$, де $\lambda_j = a + bt_j$ з невідомими (a, b) та додатними різними t_j. Перевірити гіпотезу $H_0\colon b = 0$ за критерієм відношення вірогідностей.

5. Спостерігаються дві незалежні кратні нормальні вибірки $X = (\xi_{ik}, i = \overline{1, n_k}, k = 1, 2)$, $\xi_{ik} \simeq N(\mu_k, \sigma_k^2)$ з відомими середніми та невідомими дисперсіями. Побудувати рівномірно найбільш потужний критерій відношення вірогідностей для перевірки: а) гіпотези $H_0\colon \mu_1 \le \mu_2$ проти альтернативи $H_1\colon \mu_1 > \mu_2$, у припущенні $\sigma_1^2 = \sigma_2^2$; б) гіпотези $H_0\colon \sigma_1^2 \le \sigma_2^2$ проти альтернативи $H_1\colon \sigma_1^2 > \sigma_2^2$ при відомих середніх.

6. Нехай $\xi_1, \ldots, \xi_n$ – кратна вибірка з експоненціального розподілу $\Gamma(\theta, 1)$. Побудувати критерій перевірки гіпотези H_0: $\theta = \theta_0$ проти альтернативи H_1: $\theta = \theta_1$. Знайти функцію потужності цього критерію.

7. Гіпотезу про симетричність монети перевіряють за результатами n підкидань. Довести, що критична область критерію відношення вірогідностей має вигляд $\{|\nu_n - n/2| \ge x_\alpha\}$, де ν_n – кількість аверсів.

8. Знайти рівномірно найбільш потужний критерій відношення вірогідностей для перевірки гіпотези про значення ймовірності успіху для вибірки з біноміального розподілу.

9. Знайти рівномірно найбільш потужний критерій для перевірки гіпотези $H_0\colon \theta = \theta_0$ проти альтернативи $H_1\colon \theta > \theta_0$ для вибірки з рівномірного розподілу $U(0,\theta)$.

10. Спостерігаються n випробувань Бернуллі з імовірністю успіху θ, а вибірка представлена кількістю успіхів до першого неуспіху. Знайти критерій відношення вірогідностей для перевірки гіпотези $H_0\colon \theta = \theta_0$ проти альтернативи $H_1 : \theta = \theta_1$.

11. Нехай X – кратна вибірка з нормальним розподілом $N(\mu, \sigma^2)$ спостережень та невідомим параметром $\theta = (\mu, \sigma^2)$. Побудувати найбільш потужний критерій для перевірки гіпотези $H_0\colon \mu = 0$, що спирається на статистику $T = \left(\sup\limits_{\theta\in\Theta_0} L(X,\theta)\right) \Big/ \left(\sup\limits_{\theta\in\Theta} L(X,\theta)\right)$.

12. Кратна вибірка $X = (\xi_1, \ldots, \xi_n)$ утворена спостереженнями з показниковими розподілами $Exp(\theta)$. Знайти розподіл випадкової величини τ для однобічного критерію з $c_0 = 0$.

13. Випадкові величини $(\xi_k, k \ge 1)$ незалежні та мають однакові рівномірні розподіли $U(0,\theta)$. Довести, що кількість спостережень послідовного критерію перевірки гіпотези $H_0\colon \theta = 1$ проти альтернативи $H_1 : \theta = 2$ дорівнює $\tau = \inf(k \ge 1 : \xi_k > 1)$ за умови скінченності та $\tau = n$ у протилежному випадку. Довести, що ймовірності похибок першого та другого роду дорівнюють відповідно 0 і 2^{-n}, а $\mathbf{E}(\tau \mid H_0) = n$, $\mathbf{E}(\tau \mid H_1) = 2 - 2^{1-n}$.

14. Довести, що статистика з гіпергеометричним розподілом задовольняє умову монотонності відношення вірогідностей. Побудувати відповідний найбільш потужний критерій.

15. У послідовності n випробувань Бернуллі ймовірність успіху дорівнює p. Побудувати критерій для перевірки гіпотези $H_0\colon$ $p = 0$ проти $H_1\colon p = 0.1$, у якому ймовірності похибок першого та другого роду не перевищують 0.01.

16. Нехай спостереження X за гіпотезою H_0 має щільність f_0, а за гіпотезою H_1 – f_1. Розглянемо критерій із вибірковою критичною областю $W_c = \{x \in S : f_1(x) \ge cf_0(x)\}$ для сталої $c > 0$. Позначимо через $\alpha(c), \beta(c)$ відповідні похибки першого та другого роду. Довести: а) $\beta(c)/(1-\alpha(c)) \le c \le \alpha(c)/(1-\beta(c))$, б) випадок $\alpha(c) + \beta(c) \le 1$ означає незміщеність, в) $\alpha(c) + \beta(c) \ge \alpha(1) + \beta(1)$ (це найкращий критерій при $c = 1$).

Модуль 2.
Елементи статистики випадкових процесів

2.1. Метод найменших квадратів

Література: [1, с. 423–434], [2, с. 432–441]

А

1. Дані містять кількості смертельних випадків через СНІД у Австралії протягом 12 послідовних кварталів, що починаються з другого кварталу 1983 року:

Квартал (i)	1	2	3	4	5	6	7	8	9	10	11	12
Кількість випадків (N_i)	1	2	3	1	4	9	18	23	31	20	25	37

Зобразити графік розсіяння даних. Статистик запропонував таку модель: $E[N_i] = \gamma i^2$, де γ – параметр, що буде оцінений із вищенаведених даних. Виконати таке: а) показати, що оцінка методу найменших квадратів (МНК), яка мінімізує суму $q = \sum\limits_{i=1}^{12} \left(N_i - \gamma i^2\right)^2$, дорівнює

$$\widehat{\gamma} = \left(\sum_{i=1}^{12} i^2 N_i\right) \Bigg/ \left(\sum_{i=1}^{12} i^4\right);$$

б) показати, що альтернативна зважена оцінка методу найменших квадратів, яка мінімізує $q^* = \sum\limits_{i=1}^{12} \left(N_i - \gamma i^2\right)^2 / i^2$, дорівнює

$$\widetilde{\gamma} = \left(\sum_{i=1}^{12} N_i\right) \Bigg/ \left(\sum_{i=1}^{12} i^2\right);$$

в) якщо відомо, що $\sum\limits_{i=1}^{12} i^4 = 60710$, $\sum\limits_{i=1}^{12} i^2 = 650$, обчислити $\widehat{\gamma}$ й $\widetilde{\gamma}$ для вищенаведених даних;
г) оцінити, чи є адекватною модель одного параметра в а). Розглянемо двопараметричну модель $E[N_i] = \gamma i^{\theta}$. Для оцінювання параметрів γ, θ використаємо просту лінійну модель регресії

$E[Y_i] = \alpha + \beta x_i$, де $x_i = \ln i$, $Y_i = \ln N_i$, $i = \overline{1, 12}$. Пов'язати параметри γ, θ з параметрами регресії α, β. Використовуючи значення оцінки β, провести статистичний тест для оцінювання адекватності моделі, запропонованої у пункті а).

2. Для квадратично інтегровних випадкових величин ξ, η довести, що середньоквадратичне відхилення $\mathbf{E}(\eta - a - b\xi)^2$ набуває найменшого значення при $b = \mathrm{Cov}(\xi, \eta)/\mathbf{D}\xi$, $a = \mathbf{E}\eta - b\mathbf{E}\xi$.

3. Знайти оцінку (a, b) методу найменших квадратів перпендикулярних відстаней від регресії, що мінімізує суму

$$\sum_{j=1}^{n} (\xi_j - a - bx_j)^2/(1 + b^2).$$

4. Показати, що нахил лінійної регресії, що оцінений методом найменших квадратів за трьома спостереженнями $(0, 0)$, $(1, y)$, $(2, 2)$ дорівнює 1 для всіх значень y.

B

1. Судові експерти використовують різні методи для того, щоб визначити ймовірний час смерті під час посмертної експертизи людських тіл. Недавно запропоновано об'єктивний метод, згідно з яким використано концентрації складу 3-methoxytyramine (або 3MT) у специфічній частині мозку. Для вивчення відношення між посмертним інтервалом і концентрацією 3MT були взяті зразки відповідної частини мозку, для яких час смерті визначався за показами свідка. Інтервали (x; у годинах) і концентрації (y; у частинах на мільйон) для 18 індивідуумів, що померли від органічної серцевої хвороби, задані у табл. 2.1.1. Для останніх двох індивідуумів (17-го та 18-го у таблиці) показів свідків немає, і встановлено лише інформацію про активність індивідуумів на відповідні моменти. За даними табл. 2.1.1 обчислено $\sum x = 337$, $\sum x^2 = 9854.5$, $\sum y = 42.98$, $\sum y^2 = 109.7936$, $\sum xy = 672.8$.

Таблиця 2.1.1

Номер спостереження	**Інтервал** (x)	**Концентрація** (y)
1	5.5	3.26
2	6.0	2.67
3	6.5	2.82
4	7.0	2.80
5	8.0	3.29
6	12.0	2.28
7	12.0	2.34
8	14.0	2.18
9	15.0	1.97
10	15.5	2.56
11	17.5	2.09
12	17.5	2.69
13	20.0	2.56
14	21.0	3.17
15	25.5	2.18
16	26.0	1.94
17	48.0	1.57
18	60.0	0.61

У цьому дослідженні необхідно дослідити зв'язок між концентрацією (як змінною відгуку) та інтервалом (як незалежною величиною): а) побудувати графік розсіяння (scatterplot) даних; б) обчислити коефіцієнт кореляції для даних і використати його для перевірки нульової гіпотези про те, що теоретичний коефіцієнт кореляції є нульовим; в) обчислити регресійну пряму методом найменших квадратів і використати її для оцінювання концентрації ЗМТ після закінчення 1 доби та після закінчення 2 діб; г) обчислити 99%-й вірогідний інтервал для нахилу лінії регресії. Використати цей вірогідний інтервал для перевірки гіпотези про те, що нахил дорівнює нулю.

2. Нехай $X = (\xi_j,\ j = \overline{1,n})$ – довільна вибірка така, що $\mathbf{E}\xi_j = \mu$, $\mathbf{D}\xi_j = \sigma^2$, $\mathrm{Cov}(\xi_j, \xi_k) = \rho\sigma^2$, $j \neq k$. Довести, що в класі лінійних за спостереженнями оцінок найменшу дисперсію має вибіркове середнє $\widehat{\mu}_n$.

3. Нехай $\xi_k = \theta_{l(k)} + \varepsilon_k, k = \overline{1,n}$, де $l(k) = 1$ при $k \leq m$ та $l(k) = 2$ при $k > m$, а ε_k – незалежні з розподілом $N(0, \sigma^2)$. Знайти оцінку МНК для (θ_1, θ_2).

4. Випадковий вектор (ξ, η) має сумісну щільність $f(x, y)$, а вагова функція $w(x, y) > 0$ така, що величина $(\xi^2 + \eta^2)w(\xi, \eta)$ інтегровна. Визначимо зважену оцінку МНК

$$(a^*, b^*) = \arg\min_{(a,b)\in\mathbb{R}^2} \mathbf{E}w(\xi, \eta)(\eta - a - b\xi)^2.$$

Нехай сумісна щільність вектора (ξ^*, η^*) пропорційна $f(x, y)w(x, y)$. Довести, що $b^* = \mathrm{Cov}(\xi^*, \eta^*)/\mathbf{D}\xi^*$, $a^* = \mathbf{E}\eta^* - b^*\mathbf{E}\xi^*$.

C

1. Довести, що без припущення нормальності моделі оцінка методом найменших квадратів $\widehat{\theta}$ та оцінка $\widehat{\sigma}^2_{n-k}$ є незміщеними для θ та σ^2, і $\mathrm{cov}_\theta(\widehat{\theta}) = \sigma^2 V^{-1}$.

2. Довести, що випадкові вектори $X - T'\widehat{\theta}$ і $T'(\widehat{\theta} - \theta)$ є: а) ортогональні майже напевне, б) некорельовані, в) у випадку нормальної моделі незалежні.

3. Нехай у моделі лінійної регресії $X = T'\theta + \varepsilon$ координати вектора похибок ε незалежні, а $\widehat{\theta}$ – ОМВ вектора θ. Довести: а) якщо ε_j рівномірно розподілені на інтервалі $[-\sigma, \sigma]$, то $\widehat{\theta}$ мінімізує функцію контрасту $\max\limits_{1\leq j\leq n} |\xi_j - (T'\theta)_j|$; б) якщо ε_j мають щільності Лапласа $\frac{1}{2\sigma}\exp(-|y|/\sigma)$, то значення $\widehat{\theta}$ мінімізує суму абсолютних відхилень $\sum\limits_{j=1}^{n} |\xi_j - (T'\theta)_j|$. Знайти ОМВ для σ.

4. Визначимо для випадкового вектора $\xi = (\eta, \zeta)$ змішані моменти $\mu_{ij} = \mathbf{E}\eta^i\zeta^j$ та для кратної векторної вибірки $X = (\xi_1, \ldots, \xi_n)$ вибіркові змішані моменти $\widehat{\mu}_{ij} = n^{-1}\sum\limits_{k=1}^{n} \eta_k^i \zeta_k^j$. Узагальнити метод моментів на клас змішаних моментів.

Нехай має місце модель лінійної регресії $\eta_k = a + b\zeta_k + \varepsilon_k, k = \overline{1,n}$, із некорельованими однаково розподіленими центрованими похибками ε_k. Довести, що оцінка методу моментів параметра

$\theta = (a, b)$ має вигляд

$$\widehat{b} = \left(\sum_{k=1}^{n} (\eta_k - \widehat{\eta}_n) \left(\xi_k - \widehat{\xi}_n \right) \right) / \sum_{k=1}^{n} \left(\xi_k - \widehat{\xi}_n \right)^2, \quad \widehat{a} = \widehat{\eta}_n - \widehat{b}\widehat{\xi}_n.$$

5. Нехай $(\xi_1, \ldots, \xi_n)$ – кратна вибірка з нормального розподілу на площині: $\xi_j = (\zeta_{1j}, \zeta_{2j}) \simeq N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$: а) довести, що ОМВ для $(\sigma_1^2, \sigma_2^2, \rho)$ при відомих μ_1, μ_2 матиме вигляд:

$$\widehat{\sigma}_i^2 = n^{-1} \sum_{j=1}^{n} (\zeta_{ij} - \mu_i)^2, \quad i = 1, 2,$$

$$\widehat{\rho} = \left(\sum_{j=1}^{n} (\zeta_{1j} - \mu_1)(\zeta_{2j} - \mu_2) \right) / \, n\widehat{\sigma}_1\widehat{\sigma}_2;$$

б) при невідомих μ_1, μ_2 та $n > 5$ ці оцінки визначаються як у попередній задачі.

6. Обчислити перші п'ять поліномів Чебишова.

7. Ракета рухається рівномірно та прямолінійно, спостереження її координат у моменти $t = \overline{1, 5}$ дорівнюють 22.97, 23.06, 23.33, 24.21, 23.98. Побудувати вірогідні інтервали для положення ракети у момент $t = 0$ та її швидкості на рівні 0.99.

8. Випадкові величини $(\xi_k, k = \overline{1, n})$ незалежні та мають спільну функцію розподілу вигляду $F((x - \mu)/\sigma)$ із відомою функцією розподілу F та невідомими параметрами μ, σ. Зобразивши $\xi_j = \mu + \sigma\alpha_j = \mu + \sigma m + \varepsilon_j$, де α_j мають функцію розподілу F, $m = \mathbf{E}\alpha_1$, $\varepsilon_j = \alpha_j - m$, знайти оцінки методу найменших квадратів для μ, σ.

9. Значення функції $x(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ наближено обчислені: $\xi_j = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \varepsilon_j$, $j = \overline{1, n}$, $\mathbf{E}\varepsilon_j = 0$, $\mathrm{Cov}(\varepsilon) = \sigma^2 I$. Знайти оцінки МНК $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ параметрів $\beta_0, \beta_1, \beta_2$, їх середні та коваріації. Чи є статистика $\int\limits_0^1 \left(\widehat{\beta}_0 + \widehat{\beta}_1 t + \widehat{\beta}_2 t^2 \right) dt$ незміщеною для $\int\limits_0^1 x(t) dt$?

10. Довести, що оцінка методу найменших квадратів у задачі лінійної регресії для випадку, коли коваріаційна матриця похибок $\mathrm{Cov}(\varepsilon) = \sigma^2 V$ із відомою додатно визначеною матрицею V, має вигляд $\widehat{\theta} = (TV^{-1}T')^{-1}TV^{-1}X$.

11. Спостереження ξ_j з вибірки $X = (\xi_1, \ldots, \xi_n)$ має розподіл Пуассона з параметром θx_j з відомим x_j. Знайти ОМВ для θ, обчислити її асимптотичну дисперсію. Порівняти останню з дисперсією незміщеної оцінки МНК $\left(\sum\limits_{j=1}^{n} \xi_j x_j\right) \Big/ \left(\sum\limits_{j=1}^{n} x_j^2\right)$ та пересвідчитись у неефективності ОМВ.

12. Довести, що для лінійної моделі $X = T_1'\theta_1 + T_2'\theta_2 + \varepsilon$ оцінка МНК має вигляд $\widehat{\theta}_1 = (T_1T_1')^{-1}T_1\left(X - T_2'\widehat{\theta}_2\right)$, $\widehat{\theta}_2 = (T_2VT_2')^{-1}T_2VX$, де $V = I - T_1'(T_1T_1')^{-1}T_1$.

2.2. Кореляційний та дисперсійний аналіз

Література: [1, с. 432–434], [2, с. 434–444]

А

1. Дані про смертність населення використовуються для оцінки параметрів захворюваності. Такі дані наведено для 10 регіонів у табл. 2.2.1:

Таблиця 2.2.1

Регіон	**Смертність** (m) **на 10000 осіб**	**Захворюваність** (s) **на 1000 осіб**
1	125.2	206.8
2	119.3	213.8
3	125.3	197.2
4	111.7	200.6
5	117.3	189.1
6	100.7	183.6
7	108.8	181.2
8	102.0	168.2
9	104.7	165.2
10	121.1	228.5

сумарні дані: $\sum m = 1136.1$, $\sum m^2 = 129853.03$, $\sum s = 1934.2$, $\sum s^2 = 377700.62$, $\sum ms = 221022.58$;
а) обчислити коефіцієнт кореляції між смертністю та захворюваністю та знайти P-значення гіпотези про те, що вказаний коефіцієнт дорівнює нулю. Сформулювати результат дослідження, зобразити графік залежності та пояснити характер залежності; б) обчислити оцінки коефіцієнтів регресії захворюваності відносно смертності та перевірити гіпотезу про те, що коефіцієнт нахилу не перевищує 2.0; в) перевірити гіпотезу про те, що коефіцієнт нахилу дорівнює 0; г) для регіону зі смертністю 115.0 оцінити захворюваність та обчислити 95%-і межі для цієї характеристики.

2. У дослідженні різниці в розмірах вимог між трьома різними областями отримано випадкову вибірку з незалежних вимог та проведено дисперсійний аналіз. Нижче наведено табл. 2.2.2 ANOVA з деякими пропущеними значеннями:

Таблиця 2.2.2

Джерело зміни	Ступені вільності	Сума квадратів	Середня сума квадратів
Між областями	2	4439.7	2219.9
Залишки	?	?	?
Всього	29	15153.2	

Обчислити відсутні значення в цій таблиці та провести відповідний F-тест для визначення, чи є істотна різниця між середніми значеннями вимог для цих трьох областей.

B

1. В експерименті для порівняння результативності вакцин різної сили призначили дітям 12 комплектів вакцин у 12 групах з однаковою кількістю дітей. Зафіксовано відсотки тих дітей, які залишилися здоровими після вакцинації (змінна PRH). Сила кожної партії вакцин вимірювалася у незалежних випробуваннях і реєструвалася як значення змінної SV. Отримано такі дані:

Партія	1	2	3	4	5	6
PRH(y)	16	68	23	35	42	41
SV(x)	0.9	1.6	2.3	2.7	3.0	3.3

Партія	7	8	9	10	11	12
PRH(y)	46	48	52	50	54	53
SV(x)	3.7	3.8	4.1	4.2	4.3	4.5

За ними обчислено $\sum x = 38.4$, $\sum y = 528$, $\sum x^2 = 137.16$, $\sum y^2 = 25428$, $\sum xy = 1778.4$;
а) зобразити наближений графік, щоб показати зв'язок між SV й PRH; б) обчислити повну суму квадратів, суму квадратів регресії і залишкову суму квадратів для методу найменших квадратів лінійного регресійного аналізу для змінної PRH проти SV, що використають усі 12 спостережень даних. Визначити коефіцієнт детермінації R^2. Записати рівняння лінії регресії; в) дослідити, чи дійсно на 5%-му рівні можна стверджувати, що нахил основної регресії ненульовий.

2. Нехай пряма лінія отримана методом найменших квадратів із набору даних $\{(x_i, y_i), i = \overline{1, n}\}$, що має вибірковий коефіцієнт кореляції r. Припустимо, що оцінка значення у точці x_i дорівнює $\widehat{y}_i$. Довести, що вибірковий коефіцієнт кореляції для даних $\{(y_i, \widehat{y}_i), i = \overline{1, n}\}$ також дорівнює r.

C

1. Наступні дані відносяться до спалаху ботулізму. Для кожного об'єкта дослідження – людини, що захворіла, – зареєстрований її вік у роках (змінна x), період між моментом споживання неякісної їжі та виникненням ознак хвороби (змінна y) та ознака виживання людини (Survival, змінна S). Значення $S = Y$ означає, що індивід вижив (Survival = Yes), а $S = N$ – помер (тобто, Survival = No):

Особа	01	02	03	04	05	06	07	08	09
Вік (x)	29	39	44	37	42	17	38	43	51
Період (y)	13	46	43	34	20	20	18	72	19
Ознака (S)	N	Y	Y	N	N	Y	N	Y	N

Особа	10	11	12	13	14	15	16	17	18
Вік (x)	30	32	59	33	31	32	32	36	50
Період (y)	36	48	44	21	32	86	48	28	16
Ознака (S)	N	N	Y	N	N	Y	N	Y	N

Для тих індивідів, які померли: $\sum x = 405$, $\sum y = 305$, $\sum x^2 = 15517$, $\sum y^2 = 10035$, а для тих, які вижили : $\sum x = 270$, $\sum y = 339$, $\sum x^2 = 11396$, $\sum y^2 = 19665$;
а) побудувати графік залежності періоду від віку з урахуванням ознаки; б) побудувати відповідні точкові графіки залежності між віком та виживанням, періодом та виживанням; в) побудувати 95%-ві та 99%-ві вірогідні інтервали для середніх значень змінної y окремо для групи тих індивідів, які вижили, та групи тих, що померли, а також для різниці цих величин; г) перевірити гіпотезу про те, що дисперсії періодів y для групи тих індивідів, що вижили, та групи тих, що померли, однакові.

3. Провести дисперсійний аналіз за даними табл. 2.2.3:

Таблиця 2.2.3

Джерело	**Ступені вільності**	**Сума квадратів**	**Середня сума квадратів**	F	p-**значення**
Регресія	1	1486.9	1486.9	165.69	0.000
Помилка	9	80.8	8.98		
Всього	10	1567.6			

Обчислити 95%-й вірогідний інтервал для очікуваного значення змінної PRH з умови задачі B.1 для партії вакцини з SV = 3.5

2. Розглянемо поліноміальну модель:

$$\xi_j = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 t_j^3 + \varepsilon_j,\ j \le n,$$

в якій випадкові величини $\{\varepsilon_j, j \le n\}$ – центровані, незалежні та однаково розподілені. Нехай $n = 12$, $t_j = 0$ при $j \le 4$, $t_j = -1$ при $4 < j \le 9$ та $t_j = 1$ при $9 < j \le 12$. Довести, що для наступних функцій існують незміщені оцінки: $\beta_0 + \beta_2$, β_1, $\beta_0 - \beta_1$, $\beta_1 + \beta_3$, та $\beta_0 + \beta_1 + \beta_2 + \beta_3$.

3. Величини ξ_1, ξ_2, ξ_3 некорельовані, мають додатні середні та малі коефіцієнти варіації $v_i, i = 1, 2, 3$. Довести таке наближення

для коефіцієнта кореляції:

$$\rho(\xi_1/\xi_2,\ \xi_1/\xi_3) \approx v_3^2/((v_1^2+v_3^2)(v_2^2+v_3^2))^{1/2}.$$

4. Дві двовимірні нормальні вибірки мають однакові (невідомі) коефіцієнти кореляції $\rho_1 = \rho_2 = \rho$, невідомі інші параметри та об'єми n_i, а також вибіркові коефіцієнти кореляції r_i, $i = 1, 2$. Довести, що оцінка максимальної вірогідності параметра ρ така: $\widehat{\rho} = (n(1+r_1r_2)-(n^2(1-r_1r_2)^2-4n_1n_2(r_1-r_2)^2)^{1/2})/2(n_1r_2+n_2r_1)$, де $n = n_1 + n_2$.

5. Розглянемо лінійну двофакторну модель, в якій двовимірна вибірка $X = (\xi_{ij}, i = \overline{1,n};\ j = \overline{1,m})$ задовольняє рівняння

$$\xi_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},\ i = \overline{1,n}, j = \overline{1,m},$$

де похибки $\varepsilon_{ij} \simeq N(0,\sigma^2)$ незалежні в сукупності, а вектор невідомих параметрів $\overrightarrow{\theta} = \big(\mu,\ (\alpha_i, i = \overline{1,n}),\ (\beta_j, j = \overline{1,m}),\ \sigma^2\big)$ такий, що виконуються умови центрованості: $\sum\limits_{i=1}^{n} \alpha_i = 0$, $\sum\limits_{j=1}^{m} \beta_j = 0$. Довести, що вектор

$$\widehat{\overrightarrow{\theta}} = (\overline{\xi}_{\bullet\bullet}, (\overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet\bullet}, i = \overline{1,n}), (\overline{\xi}_{\bullet j} - \overline{\xi}_{\bullet\bullet}, j = \overline{1,m}), \widehat{\sigma}^2)$$

є оптимальною оцінкою для $\overrightarrow{\theta}$, що складається з чотирьох незалежних груп. Тут групові середні статистики мають вигляд

$$\overline{\xi}_{\bullet\bullet} = \sum_{i=1}^{n}\sum_{j=1}^{m} \xi_{ij}/nm,\ \overline{\xi}_{i\bullet} = \sum_{j=1}^{m} \xi_{ij}/m,\quad \overline{\xi}_{\bullet j} = \sum_{i=1}^{n} \xi_{ij}/n,$$

$$\widehat{\sigma}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} (\xi_{ij} - \overline{\xi}_{i\bullet} - \overline{\xi}_{\bullet j} + \overline{\xi}_{\bullet\bullet})^2/(n-1)(m-1).$$

2.3. Оптимальний прогноз для стаціонарних послідовностей

Література: [1, с. 435–457], [2, с. 445–469]

A

1. Довести, що вінерівський процес $(w(t), t \in [0, a])$ є процесом з ортогональними приростами. Знайти його спектральну функцію.

2. Нехай $(\eta_n, n \geq 0)$ незалежні однаково розподілені дійсні випадкові величини, а $(\nu(t), t \geq 0)$ – незалежний від них процес Пуассона. Довести, що випадковий процес $\xi_t = \exp\left(i\eta_{\nu(t)}\right)$ є стаціонарним другого порядку, та знайти його коваріаційну функцію.

3. Знайти оптимальний прогноз для $\widehat{\xi}_1$, якщо спектральна щільність дорівнює $f(\lambda) = (25 + 24\cos\lambda)^{-1}$.

B

1. Довести, що для квадратично сумованої коваріаційної функції: $\sum\limits_{t\in\mathbb{Z}} |r(t)|^2 < \infty$ відповідна спектральна функція F має щільність вигляду $f(\lambda) = (2\pi)^{-1} \sum\limits_{t\in\mathbb{Z}} \exp(-it\lambda) r(t)$.

2. Довести, що при $\alpha > 0$ випадковий процес

$$\xi_t = \exp(-\alpha t) W(\exp(2\alpha t)), t \in \mathbb{R}_+,$$

є стаціонарним другого порядку (процесом Орнштейна – Уленбека), та знайти його коваріаційну функцію. Тут $W(s)$ – вінерівський процес.

3. Знайти оптимальний прогноз для $\widehat{\xi}_1$, якщо спектральна щільність $f(\lambda) = 5 + 4\cos\lambda$.

C

1. Випадкові величини $(\varepsilon_n, n \in \mathbb{Z})$ такі, що $\mathbf{E}\varepsilon_n = 0$, $\mathbf{E}\varepsilon_n\overline{\varepsilon_k} = \delta_{nk}$, а $(c_k, k = \overline{0,m})$ – довільні сталі. Тоді послідовність $\xi_n = \sum\limits_{k=0}^{m} c_k \varepsilon_{n-k}$ є стаціонарною другого порядку. Знайти її коваріаційну функцію.

2. Строго стаціонарна послідовність $(\xi_n, n \in \mathbb{Z})$ є центрованою: $\mathbf{E}\xi_n = 0$, та має стохастичний спектральний процес $(\zeta(\lambda))$. Довести, що має місце збіжність: $n^{-1}\sum\limits_{k=0}^{n-1}\xi_n \xrightarrow{L_2} \widetilde{\zeta}(\{0\}), n \to \infty$, де $\widetilde{\zeta}$ – відповідна стохастична міра.

3. Строго стаціонарна послідовність $(\xi_n, n \in \mathbb{Z})$ така, що виконуються умови $\mathbf{E}\xi_n = 0$ та $\mathbf{E}\left|\sum\limits_{k=0}^{n-1}\xi_k\right|^2 = O\left(n^{2-\delta}\right), n \to \infty$, при деякому $\delta > 0$. Довести, що $n^{-1}\sum\limits_{k=0}^{n-1}\xi_k \xrightarrow{P1} 0, n \to \infty$.

4. Розклад Карунена – Лоєва – Пугачова.

Нехай $\xi(t), t \in [0,1]$, с.к. неперервний дійсний випадковий процес із нульовим середнім та коваріаційною функцією $r(t,s) = \mathbf{E}\xi(t)\xi(s)$. Ця функція неперервна та інтегровна з квадратом на $[0,1]^2$. Тому інтегральне рівняння Фредгольма у просторі $L_2[0,1]$:

$$\varphi(t) = \lambda\int\limits_0^1 r(t,s)\varphi(s)ds$$

має для деяких $\lambda = \lambda_n > 0$ розв'язки $\varphi_n \in L_2[0,1], n \geq 1$, що утворюють ортонормований базис. Зокрема, справедливе зображення

$$r(t,s) = \sum_{n=1}^{\infty}\lambda_n^{-1}\varphi_n(t)\varphi_n(s).$$

Суму праворуч можна розглядати як спектральний розклад за лічильною мірою m, що має одиничні стрибки на $\mathbb{N}$. Застосувавши до цього розкладу теорему Карунена, побудуйте випадковий процес $\zeta(t)$ на $[0,1]$ з ортогональними приростами та зі спектральною функцією m. Позначимо $\zeta_n = \zeta(n+0) - \zeta(n-0)$ величини ненульових стрибків цього процесу. Тоді з теореми Карунена отримуємо зображення

$$\xi(t) = \sum_{n=1}^{\infty}\lambda_n^{-1/2}\varphi_n(t)\ \zeta_n$$

з попарно ортогональними та нормованими $\zeta_n \in L_2(\mathbf{P})$.

Зокрема, для вінерівського процесу $w(t)$ маємо $r(t,s) = \min(t,s)$, $\lambda_n = \pi^2(n-1/2)^2$, $\varphi_n(t) = \sqrt{2}\sin(n-1/2)\pi t$, та

$$w(t) = \sqrt{2}\sum_{n=1}^{\infty} \zeta_n\ \pi^{-1}(n-1/2)^{-1}\sin(n-1/2)\pi t.$$

5. Нехай (ζ_n) – незалежні в сукупності стандартні нормальні величини. Довести: а) якщо $(\varphi_n(t), n \geq 1)$ – ортонормований базис у $L^2([0,1])$, то випадковий процес $w(t) = \sum_{n\geq 1} \zeta_n \int_0^t \varphi_n(s)ds$ є вінерівським процесом; б) сума збіжного у середньоквадратичному на $[0, 2\pi)$ ряду $w(t) = \sum_{n\in\mathbb{Z}} \zeta_n(1-\exp(int))/in$ є вінерівським процесом на $[0, 2\pi)$ (при $n = 0$ коефіцієнт визначається за неперервністю).

6. Нехай (ξ_n) – стаціонарна послідовність, випадкова величина $\eta \in H_2(\xi)$ має спектральну характеристику $g(\lambda)$, а θ – оператор зсуву. Довести, що спектральна характеристика $\theta\eta$ дорівнює $\exp(i\lambda)g(\lambda)$.

ДОДАТКИ

Додаток 1. Нормальний розподіл

У табл. 1 наведено значення функції $\Phi(t)$ нормального розподілу з параметрами $(0;\,1)$:

- для заданих $t \geq 0$ табульовані значення функції

$$N_{0;\,1}\left(t\right) = \Phi\left(t\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} \exp\left\{-\frac{s^2}{2}\right\} ds.$$

- для $t < 0$ значення функції $\Phi(t)$ отримуються з рівності

$$\Phi(t) = 1 - \Phi(-t).$$

Значення $N_{a;\,\sigma^2}\left(x\right)$ функції нормального розподілу з параметрами a і σ^2 обчислюється за значеннями табульованої функції $N_{0;\,1}\left(t\right) = \Phi\left(t\right)$ нормального розподілу $N_{0;\,1}$:

$$N_{a;\,\sigma^2}\left(x\right) = N_{0;\,1}\left(\frac{x-a}{\sigma}\right) = \Phi\left(\frac{x-a}{\sigma}\right).$$

Табл. 1 допускає лінійну інтерполяцію.

Таблиця 1. **Значення функції** $\Phi(t)$

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9900	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	9984	,9985	,9985	,9986	,9986
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	,9987	,9990	,9993	,9995	,9997	,9998	,9998	,9999	,9999	1,000

Додаток 2. Розподіл хі-квадрат Пірсона

У табл. 2 наведено значення функції $\chi^2_{\alpha;n}$, що задає квантилі рівня $1-\alpha$ розподілу Пірсона (χ^2- розподілу) із n ступенями вільності. За означенням, $P(\chi^2_n > \chi^2_{\alpha;n}) = \alpha$.

Таблиця 2. **Значення функції** $\chi^2_{\alpha;n}$

n	Значення α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
1	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,64
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21
3	0,12	0,22	0,35	0,58	6,23	7,82	9,35	11,34
4	0,30	0,48	0,71	1,06	7,78	9,48	11,14	13,28
5	0,55	0,83	1,14	1,61	9,24	11,07	12,83	15,09
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09
9	2,90	2,70	3,32	4,17	14,68	16,92	19,02	21,67
10	2,56	3,25	3,94	4,86	15,99	18,31	20,48	23,21
11	3,05	3,82	4,58	5,58	17,28	19,68	21,92	24,72
12	3,57	4,40	5,23	6,3	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41
18	7,02	8,23	9,39	10,86	25,99	28,87	31,53	34,81
19	7,63	8,91	10,12	11,65	27,2	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57

Закінчення табл. 2

n	Значення α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,92
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,26
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89
40	22,16	24,43	26,51	29,05	51,80	55,76	59,34	63,69
50	29,71	32,36	34,76	37,69	63,17	67,50	71,42	76,15
60	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38
70	45,44	48,76	51,74	55,33	85,53	90,53	95,02	100,4
80	53,54	57,15	60,39	64,28	96,58	101,9	106,6	112,3
90	61,75	65,65	69,13	73,29	107,6	113,1	118,1	124,1
100	70,06	74,22	77,93	82,36	118,5	124,3	129,6	135,8

Додаток 3. Розподіл Стьюдента

У табл. 3 наведено значення функції $t_{\alpha;n}$, що задає квантилі рівня $1 - \alpha$ розподілу Стьюдента (t-розподілу) із n ступенями вільності. За означенням, $P(t_n > t_{\alpha;n}) = \alpha$.

Таблиця 3. **Значення функції** $t_{\alpha;\,n}$

n	Значення α					n	Значення α			
	0,050	0,025	0,010	0,005			0,050	0,025	0,010	0,005
1	6,314	12,706	31,821	63,657		18	1,734	2,101	2,552	2,878
2	2,920	4,303	6,965	9,925		19	1,729	2,093	2,539	2,861
3	2,353	3,182	4,541	5,841		20	1,725	2,086	2,528	2,845
4	2,132	2,776	3,747	4,604		21	1,721	2,080	2,518	2,831
5	2,015	2,571	3,365	4,032		22	1,717	2,074	2,508	2,819
6	1,943	2,447	3,143	3,707		23	1,714	2,069	2,500	2,807
7	1,895	2,365	2,998	3,499		24	1,711	2,064	2,492	2,797
8	1,860	2,306	2,896	3,355		25	1,708	2,060	2,485	2,787
9	1,833	2,262	2,821	3,250		26	1,706	2,056	2,479	2,779
10	1,812	2,228	2,764	3,169		27	1,703	2,052	2,473	2,771
11	1,796	2,201	2,718	3,106		28	1,701	2,048	2,467	2,763
12	1,782	2,179	2,681	3,055		29	1,699	2,045	2,462	2,756
13	1,771	2,160	2,651	3,012		30	1,697	2,042	2,457	2,750
14	1,761	2,145	2,624	2,977		40	1,684	2,021	2,423	2,704
15	1,753	2,131	2,602	2,947		60	1,671	2,000	2,390	2,660
16	1,746	2,120	2,583	2,921		120	1,658	1,980	2,358	2,617
17	1,740	2,110	2,567	2,898		∞	1,645	1,960	2,326	2,576

Додаток 4. Розподіл Фішера

У таблицях 4.1 та 4.2 наведено значення функції $F_{\alpha;n;m}$, що задає квантилі рівня $1-\alpha$ для функції розподілу Фішера (F-розподілу) із n, m ступенями свободи, при вірогідних рівнях $\alpha = 0,05$ та $\alpha = 0,01$. За означенням, $P(F_{n;m} > F_{\alpha;n;m}) = \alpha$.

Таблиця 4.1. **Значення функції $F_{\alpha;n;m}$ (вірогідний рівень 0,05)**

m	n									
	1	2	3	4	5	6	7	8	9	10
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,57
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93

Закінчення табл. 4.1

m	n									
	12	14	16	18	20	30	40	50	60	100
3	8,74	8,71	8,69	8,67	8,66	8,62	8,59	8,58	8,57	8,55
4	5,91	5,87	5,84	5,82	5,80	5,75	5,72	5,70	5,69	5,66
5	4,68	4,64	4,60	4,58	4,56	4,50	4,46	4,44	4,43	4,41
6	4,00	3,96	3,92	3,90	3,87	3,81	3,77	3,75	3,74	3,71
7	3,57	3,53	3,49	3,47	3,44	3,38	3,34	3,32	3,30	3,27
8	3,28	3,24	3,20	3,17	3,15	3,08	3,04	3,02	3,01	2,97
9	3,07	3,03	2,99	2,96	2,94	2,86	2,83	2,80	2,79	2,76
10	2,91	2,86	2,83	2,80	2,77	2,70	2,66	2,64	2,62	2,59
11	2,79	2,74	2,70	2,67	2,65	2,57	2,53	2,51	2,49	2,46
12	2,69	2,64	2,60	2,57	2,54	2,47	2,43	2,40	2,38	2,35
13	2,60	2,55	2,51	2,48	2,46	2,38	2,34	2,31	2,30	2,26
14	2,53	2,48	2,44	2,41	2,39	2,31	2,27	2,24	2,22	2,19
15	2,48	2,42	2,38	2,35	2,33	2,25	2,20	2,18	2,16	2,12
16	2,42	2,37	2,33	2,30	2,28	2,19	2,15	2,12	2,11	2,07
17	2,38	2,33	2,29	2,26	2,23	2,15	2,10	2,08	2,06	2,02
18	2,34	2,29	2,25	2,22	2,19	2,11	2,06	2,04	2,02	1,98
19	2,31	2,26	2,21	2,18	2,16	2,07	2,03	2,00	1,98	1,94
20	2,28	2,22	2,18	2,15	2,12	2,04	1,99	1,97	1,95	1,91
22	2,23	2,17	2,13	2,10	2,07	1,98	1,94	1,91	1,89	1,85
24	2,18	2,13	2,09	2,05	2,03	1,94	1,89	1,86	1,84	1,80
26	2,15	2,09	2,05	2,02	1,99	1,90	1,85	1,82	1,80	1,76
28	2,12	2,06	2,02	1,99	1,96	1,87	1,82	1,79	1,77	1,73
30	2,09	2,04	1,99	1,96	1,93	1,84	1,79	1,76	1,74	1,70
40	2,00	1,95	1,90	1,87	1,84	1,74	1,69	1,66	1,64	1,59
50	1,95	1,89	1,85	1,81	1,78	1,69	1,63	1,60	1,58	1,52
60	1,92	1,86	1,82	1,78	1,75	1,65	1,59	1,56	1,53	1,48
100	1,85	1,79	1,75	1,71	1,68	1,57	1,52	1,48	1,45	1,39

Таблиця 4.2. **Значення функції $F_{\alpha;n;m}$ (вірогідний рівень 0,01)**

m	n									
	1	2	3	4	5	6	7	8	9	10
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50

Закінчення табл. 4.2

m	n									
	12	14	16	18	20	30	40	50	60	100
3	27,1	26,9	26,8	26,8	26,7	26,5	26,4	26,4	26,3	26,2
4	14,4	14,2	14,2	14,1	14,0	13,8	13,7	13,7	13,7	13,6
5	9,89	9,77	9,68	9,61	9,55	9,38	9,29	9,24	9,20	9,13
6	7,72	7,60	7,52	7,45	7,40	7,23	7,14	7,09	7,06	6,99
7	6,47	6,36	6,27	6,21	6,16	5,99	5,91	5,86	5,82	5,75
8	5,67	5,56	5,84	5,41	5,36	5,20	5,12	5,07	5,03	4,96
9	5,11	5,00	4,92	4,86	4,81	4,65	4,57	4,52	4,48	4,42
10	4,71	4,60	4,52	4,46	4,41	4,25	4,17	4,12	4,08	4,01
11	4,40	4,29	4,21	4,15	4,10	3,94	3,86	3,81	3,78	3,71
12	4,16	4,05	4,97	3,91	3,86	3,70	3,62	3,57	3,54	3,47
13	3,96	3,86	3,78	3,72	3,66	3,51	3,43	3,38	3,34	3,27
14	3,80	3,70	3,62	3,56	3,51	3,35	3,27	3,22	3,18	3,11
15	3,67	3,56	3,49	3,42	3,37	3,21	3,13	3,08	3,05	2,98
16	3,55	3,45	3,37	3,31	3,26	3,10	3,02	2,97	2,93	2,86
17	3,46	3,35	3,27	3,21	3,16	3,00	2,92	2,87	2,83	2,76
18	3,37	3,27	3,19	3,13	3,08	2,92	2,84	2,78	2,75	2,68
19	3,30	3,19	3,12	3,05	3,00	2,84	2,76	2,71	2,67	2,60
20	3,23	3,13	3,05	2,99	2,94	2,78	2,69	2,64	2,61	2,54
22	3,12	3,02	2,94	2,88	2,83	2,67	2,58	2,53	2,50	2,42
24	3,03	2,93	2,85	2,79	2,74	2,58	2,49	2,44	2,40	2,33
26	2,96	2,86	2,78	2,72	2,66	2,50	2,42	2,36	2,33	2,25
28	2,90	2,79	2,72	2,65	2,60	2,44	2,35	2,30	2,26	2,19
30	2,84	2,74	2,66	2,60	2,55	2,39	2,30	2,25	2,21	2,13
40	2,66	2,56	2,48	2,42	2,37	2,20	2,11	2,06	2,02	1,94
50	2,56	2,46	2,38	2,32	2,27	2,10	2,01	1,95	1,91	1,82
60	2,50	2,39	2,31	2,25	2,20	2,03	1,94	1,88	1,84	1,75
100	2,37	2,26	2,19	2,12	2,07	1,89	1,80	1,73	1,69	1,60

Додаток 5. Критерій Колмогорова

У табл. 5 наведено критичні значення $\varepsilon_{\alpha;n}$ – верхньої межі модуля різниці істинної та емпіричної функцій розподілів.

Значення $\varepsilon_{\alpha;n}$ для заданих α та n визначається як мінімальне ε, для якого

$$P\left\{\sup_x \left|F(x) - \hat{F}_n(x)\right| \geq \varepsilon\right\} \leq \alpha.$$

Таблиця 5. **Критичні значення** $\varepsilon_{\alpha;n}$

n	Значення α				n	Значення α		
	0,05	0,02	0,01			0,05	0,02	0,01
1	0,9750	0,9900	0,9950		25	0,2640	0,2952	0,3166
2	0,8419	0,9000	0,9293		30	0,2417	0,2702	0,2899
3	0,7076	0,7846	0,8290		35	0,2243	0,2507	0,2690
4	0,6239	0,6889	0,7342		40	0,2101	0,2349	0,2520
5	0,5633	0,6272	0,6685		45	0,1984	0,2218	0,2380
6	0,5193	0,5774	0,6166		50	0,1884	0,2107	0,2260
7	0,4834	0,5384	0,5758		55	0,1798	0,2011	0,2157
8	0,4543	0,5065	0,5418		60	0,1723	0,1927	0,2067
9	0,4300	0,4796	0,5133		65	0,1657	0,1853	0,1988
10	0,4093	0,4566	0,4889		70	0,1598	0,1796	0,1917
11	0,3912	0,4367	0,4677		75	0,1544	0,1727	0,1853
12	0,3754	0,4192	0,4491		80	0,1496	0,1673	0,1795
13	0,3614	0,4036	0,4325		85	0,1452	0,1624	0,1742
14	0,3489	0,3897	0,4176		90	0,1412	0,1579	0,1694
15	0,3376	0,3771	0,4042		95	0,1375	0,1537	0,1649
20	0,2941	0,3287	0,3524		100	0,1340	0,1499	0,1608

При $n > 100$ слід користуватися асимптотичними границями $\varepsilon_{0{,}05;n} = \frac{1{,}36}{\sqrt{n}}$; $\varepsilon_{0{,}01;n} = \frac{1{,}63}{\sqrt{n}}$, для яких справжні коефіцієнти надійності навіть трохи більші від 0,95 і 0,99 відповідно.

Додаток 6. Критерій Вілкоксона

У табл. 6 наведено нижні критичні значення $W_{\alpha;n;m}$ статистики Вілкоксона S_{nm}. Значення $W_{\alpha;n;m}$ для заданих α (вірогідного рівня) та n і m – об'ємів вибірок визначається як найбільше ціле t, для якого $P\{S_{nm} \le t\} \le \alpha$. Рівень $x_{nm}(\alpha)$, що визначається умовою $P(S_{nm} \ge x_{nm}(\alpha)) \approx \alpha$, виражається через $W_{\alpha;n;m}$ таким чином:
$x_{nm}(\alpha) = mn + \frac{n(n+1)}{2} - \left(W_{\alpha;n;m} - \frac{n(n+1)}{2}\right) = n(n+m+1) - W_{\alpha;n;m}$.

При значеннях n та m більших, ніж наведені в табл. 6 (а фактично при n і m, які задовольняють нерівності $\min\{n,\ m\} \ge 6$, $m + n \ge 20$), використовується нормальне наближення

$$x_{nm}(\alpha) = n(n+m+1)/2 + x_\alpha \sqrt{nm(n+m+1)/12},$$

де x_α – квантиль рівня $1-\alpha$ стандартного нормального розподілу.

Для побудови двосторонніх критеріїв слід мати на увазі, що за нульової гіпотези статистика S_{nm} розподілена симетрично відносно значення $n(n+m+1)/2$.

Таблиця 6. **Нижні критичні значення $W_{\alpha;\, n;\, m}$ статистики Вілкоксона**

Об'єм		Значення α					Об'єм		Значення α			
n	m	0,005	0,01	0,025	0,05		n	m	0,005	0,01	0,025	0,05
6	6	23	24	26	28		6	18	37	40	45	49
	7	24	25	27	29			19	38	41	46	51
	8	25	27	29	31		7	7	32	34	36	39
	9	26	28	31	33			8	34	35	38	41
	10	27	29	32	35			9	35	37	40	43
	11	28	30	34	37			10	37	39	42	45
	12	30	32	35	38			11	38	40	44	47
	13	31	33	37	40			12	40	42	46	49
	14	32	34	38	42			13	41	44	48	52
	15	33	36	40	44			14	43	45	50	54
	16	34	37	42	46			15	44	47	52	56
	17	36	39	43	47			16	46	49	54	58

Закінчення табл. 6

Об'єм		Значення α					Об'єм		Значення α			
n	m	0,005	0,01	0,025	0,05		n	m	0,005	0,01	0,025	0,05
7	17	47	51	56	61		9	13	65	68	73	78
	18	49	52	58	63			14	67	71	76	81
8	8	43	45	49	51			15	69	73	79	84
	9	45	47	51	54			16	72	76	82	87
	10	47	49	53	56		10	10	71	74	78	82
	11	49	51	55	59			11	73	77	81	86
	12	51	53	58	62			12	76	79	84	89
	13	53	56	60	64			13	79	82	88	92
	14	54	58	62	67			14	81	85	91	96
	15	56	60	65	69			15	84	88	94	99
	16	58	62	67	72		11	11	87	91	96	100
	17	60	64	70	75			12	90	94	99	104
9	9	56	59	62	66			13	93	97	103	108
	10	58	61	65	69			14	96	100	106	112
	11	61	63	68	72		12	12	105	109	115	120
	12	63	66	71	75			13	109	113	119	125

Модульний контроль за семестр

Модуль 1. Статистичне оцінювання і перевірка гіпотез	**0 – 50 балів**
• робота на аудиторних заняттях	0 – 2 бали
• виконання позааудиторних робіт	0 – 28 балів
• модульна контрольна робота	0 – 20 балів
Модуль 2. Елементи статистики випадкових процесів	**0 – 10 балів**
• робота на аудиторних заняттях	0 – 1 бал
• виконання позааудиторних робіт	0 – 9 балів
Іспит	**0 – 40 балів**
Всього	**0 – 100 балів**

Список рекомендованої літератури

1. *Карташов М. В.* Теорія ймовірностей і математична статистика / М. В. Карташов – К. : ВПЦ Київський університет, 2009.
2. *Карташов М. В.* Імовірність, процеси, статистика / М. В. Карташов – К. : ВПЦ Київський університет, 2006.
3. *Ивченко Г. И.* Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – М. : Высш. шк., 1964.
4. *Скороход А. В.* Елементи теорії ймовірностей та випадкові процеси / А. В. Скороход. – К. : Вища шк., 1975.
5. *Гихман И. И.* Теория вероятностей и математическая статистика / И. И. Гихман, А. В. Скороход, М. И. Ядренко. – К. : Вища шк., 1988.

Навчальне видання

МЕТОДИЧНІ ВКАЗІВКИ
ДО ЛАБОРАТОРНИХ ТА САМОСТІЙНИХ РОБІТ ІЗ ДИСЦИПЛІНИ
"МАТЕМАТИЧНА СТАТИСТИКА"

ДЛЯ СТУДЕНТІВ МЕХАНІКО-МАТЕМАТИЧНОГО ФАКУЛЬТЕТУ

Упорядники:
О. І. Василик, М. В. Карташов, В. П. Кнопова,
К. В. Ральченко, А. Ю. Рижов, Г. М. Шевченко

Редактор Л. В. Магда

Оригінал-макет виготовлено Видавничо-поліграфічним центром "Київський університет"

Формат $60x84^{1/16}$. Ум. друк. арк. 4,88. Наклад 100. Зам. № 214-6996.
Вид. № М7. Гарнітура Times New Roman. Папір офсетний. Друк офсетний.
Підписано до друку 15.04.14

Видавець і виготовлювач
Видавничо-поліграфічний центр "Київський університет",
б-р Т. Шевченка, 14, м. Київ, 01601
☎ (38044) 239 32 22; (38044) 239 31 72; тел./факс (38044) 239 31 28
e-mail: vpc@univ.kiev.ua
http: vpc.univ.kiev.ua
Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02