МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Методичні вказівки

до практичних робіт з дисципліни
"**Статистичний аналіз океанологічних процесів**"
Для студентів-бакалаврів очної форми навчання V курсу
гідрометеорологічного факультету
Напрям підготовки – гідрометеорологія
Спеціальність – океанологія

"Затверджено"
на засіданні методичної комісії
гідрометеорологічного інституту
Протокол № 2 від " 25" 10 2010р.
Голова комісії______ Єхніч М. П.

"Затверджено"
на засіданні кафедри океанології
та морського природокористування
Протокол № ___від " " ____2010р.
Зав. каф._______д.г.н. Михайлов В.І

ОДЕСА 2010

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Методичні вказівки
до практичних робіт з дисципліни
“Статистичний аналіз океанологічних процесів”

ОДЕСА 2010

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Методичні вказівки
до практичних робіт з дисципліни
“Статистичний аналіз океанологічних процесів”

“Затверджено”
на засіданні методичної ради
гідрометеорологічного інституту
Протокол № ___від “ “ ___2010р.

ОДЕСА 2010

Методичні вказівки до практичних робіт з дисципліни “Статистичний аналіз океанологічних процесів” для студентів очної форми навчання V курсу спеціальності “Океанологія” /укладач: доц. Ілюшин Віктор Якович, Одеса, ОДЕКУ, 2010 р., 35 с., укр. мова.

Вступ

Для океанологів головний інтерес представляють гідрологічні процеси, що розвиваються в часі і просторі у вигляді певної послідовності подій. Тимчасові ряди можуть бути дискретними або безперервними і володіти деякими важливими властивостями, з погляду можливості їх обробки статистичними методами. До основних властивостей, в першу чергу, відносяться ***стаціонарність***, присутність катастрофічних подій, трендів і періодичних складових процесу, їх нормальність.

Стаціонарність процесу грає важливу роль. Методи аналізу нестаціонарних процесів істотно громіздкі. Катастрофічна подія повинна бути виключена із звичайної реалізації випадкового процесу, як така, що повторюється надзвичайно рідко (надалі зазвичай піддається детермінованому аналізу). Якщо встановлено, що в процесі містяться періодичні складові, то це дозволяє уникнути певних помилок при інтерпретації результатів аналізу. ***Припущення про нормальність*** дозволяє істотно спростити аналітичне дослідження властивостей випадкового процесу (що не містять тренду, періодичних складових і катастрофічних явищ), тому перевірка цієї гіпотези завжди бажана.

Нижче розглянуті методи, що стосуються оцінювання цієї властивості. Проте передуватимемо цьому викладу деякими зауваженнями.

Мабуть, найбільш простій спосіб оцінювання стаціонарності реалізації полягає в розгляді фізичної природи процесу, якому ця реалізація належить. Якщо основні фізичні чинники, що визначають процес, не залежать від часу, то можна без подальшого дослідження вважати процес, що вивчається, стаціонарним. На практиці не всі фізичні сторони процесу чітко ясні. У подібних випадках гіпотеза про стаціонарність повинна бути перевірена шляхом аналізу наявних реалізацій. Способи перевірки можуть бути різними, від візуального перегляду реалізацій до детального статистичного оцінювання різних параметрів процесу. В усякому разі, при цьому завжди робляться істотні допущення.

По-перше, допускається, що будь-яка реалізація правильно відображає нестаціонарний характер процесу, що вивчається.

По-друге, допускається також, що довжина даної реалізації істотно більше періоду самої низькочастотної складової процесу.

По - третє, за аналізом припускається ***аналогія між частотою та вірогідністю***. Це припущення частіше за все не враховують, хоча це положення відомо з початку досліджень вірогідних процесів (за часів Лапласа, Ньютона и т.і.).

Оцінювання основних властивостей процесу здійснюється як складова частина загального аналізу, виконується до початку детального аналізу.

Завдання виконуються по вибору, залежно від рівня підготовки студента і часу відведеного йому на аудиторну і самостійну роботу по даній дисципліні.

Завдання 1

АНАЛІЗ І ОПИС ТЕНДЕНЦІЙ В РОЗВИТКУ ОКЕАНОЛОГІЧНИХ ПРОЦЕСІВ

1. Перевірка гіпотези про наявність тренда середнього і дисперсії в даних спостережень.

1.1 Основні поняття.

1.1.1 Перевірка способом різниці середніх

Цей спосіб розроблений для малих вибірок; при цьому передбачається, що вони мають нормальний розподіл.

Отже, початковий ряд розбивається на дві частини. Значення кожною з них розглядаються як дві вибірки. Перша має середню $\overline{x}_1$, друга - $\overline{x}_2$. Необхідно перевірити гіпотезу про істотність різниці $\overline{x}_1$-$\overline{x}_2$. Випробування різниці середніх, обчислених для кожної з цих сукупностей, покаже, чи істотно розрізняються між собою середні або цю розбіжність можна приписати дії випадковості і, таким чином, укласти, що тренд відсутній. Перевірка даної гіпотези спирається на t - статистику Стьюдента.

Порядок виконання. У найпростішому випадку, при рівності або неістотній відмінності дисперсій двох досліджуваних реалізацій ($\sigma_1^2 \cong \sigma_2^2$), t - статистика обчислюється, як відомо, по формулі

$$\hat{t} = \frac{\overline{x}_1 - \overline{x}_2}{\hat{\sigma}_{\bar{x}}}, \tag{1.1}$$

де $\overline{x}_1$ і $\overline{x}_2$ - середні для першої і другої сукупностей спостережень (у нашому випадку – для першої і другої половини ряду);

n_1 і n_2 - кількості спостережень в цих групах;

$\hat{\sigma}_{\bar{x}}$ - вибіркове середнє квадратичне відхилення різниці середніх.

Необхідне значення $\hat{\sigma}_{\bar{x}}$ можна визначити на основі середніх зважених величин дисперсій окремих сукупностей реалізації:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{(n_1-1)^2\,\hat{\sigma}_1^2 + (n_2-1)^2\,\hat{\sigma}_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \tag{1.2}$$

Відмітимо, що при оцінюванні дисперсій для першої $\hat{\sigma}_1^2$ і другої $\hat{\sigma}_2^2$ сукупностей береться число ступеня свободи, що дорівнює (n_1-1) і (n_2-1) відповідно.

Гіпотеза про відсутність тренда (нульова гіпотеза) відкидається при $\hat{t} \ge t_{\frac{\alpha}{2}}$ і приймається при $\hat{t} < t_{\frac{\alpha}{2}}$ Тут $\hat{t}$ - розрахункове значення t -

статистики Стьюдента, $t_{\frac{\alpha}{2}}$ - табличне її значення при рівні значущості α. Значення $t_{\frac{\alpha}{2}}$ береться з числом ступеня свободи, рівним $(n_1 + n_2 - 2)$ з табл. 4.3 (див. додаток).

Вище було відмічено, що при використанні формули (1.1) передбачається, що дисперсія двох сукупностей трохи розрізняються між собою. Перевірка однорідності дисперсій реалізується з допомогою F - критерію Фішера, який заснований на порівнянні розрахункового відношення з табличним

$$\hat{F} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}, \tag{1.3}$$

де $\hat{\sigma}_1^2 > \hat{\sigma}_2^2$.

Якщо розрахункове значення $\hat{F}$ менше, ніж табличне $F_{n_1;n_2;\alpha}$, при заданому рівні значущості (див. табл. 4.4, додаток), то можна прийняти гіпотезу про рівність дисперсій. Якщо ж $\hat{F}$ більше, ніж значення табличне $F_{n_1;n_2;\alpha}$, то гіпотеза про рівність дисперсій відхиляється і, отож, формула (1.1) для випробування різниці середніх неприйнятна.

Розглянутий метод у ряді випадків дає цілком прийнятні результати. Проте слід зазначити, що йому властиві вельми істотні дефекти. Перш за все, він застосовний тільки для рядів з монотонною тенденцією. Якщо ж ряд загальну тенденцію міняє, то точка повороту тренда може виявитися близькою до середини ряду. Через це середні двох сукупностей ряду будуть близькі, і перевірка може не показати наявність тренда. Разом з цим можна висунути і серйозніше заперечення, засноване на тому, що величина середнього квадратичного відхилення, по якій проводиться формування різниці середніх в (1.1), в динамічному ряду залежна не тільки від мінливості рівнів характеристик, але і від самого тренда. Інакше кажучи, існування тренда позначається на показнику середнього квадратичного відхилення.

Перевірка "способом різниці середніх" не є єдиним засобом перевірки гіпотез про наявність тренда середнього динамічного ряду. Для цього запропонований ряд достатньо простих критеріїв, заснованих на різних підходах – використанні поворотних точок, кореляції рангів і ін. Розглянемо один з найпростіших методів, запропонований Ф. Фостером і А. Стюартом, що дає практично надійніші результати, ніж останні. Цей метод цінний тим, що дозволяє також виявляти тренд дисперсії.

Спрощений, але прийнятний в практичній роботі підхід до проблеми оцінки різниці середніх при $n_1 \neq n_2$ і неоднорідних дисперсіях використовується в роботі Дж. Снедекора.

1.1.2 Метод Фостера – Стюарта

Ф.Фостер і А.Стюарт запропонували за даними досліджуваного ряду визначати величини u_t і l_t. Значення u_t і l_t знаходяться шляхом послідовного порівняння рівнів змінної $x(t)$. Якщо яке-небудь значення ряду перевищує по своїй величині кожне з попередніх значень, то величині u_t присвоюється значення 1, в решті випадків - 0.

Таким чином

$$u_t = \begin{cases} 1, \text{ якщо } \hat{x}_t > x_{t-1}, x_{t-2}, ..x_1, \\ 0, \text{ в решті випадків.} \end{cases} \quad (1.4)$$

Навпаки, якщо рівень менше всіх попередніх, то l_t присвоюють значення 1. таким чином

$$l_t = \begin{cases} 1, \text{ якщо } x_t < x_{t-1}, x_{t-2}, ..x_1, \\ 0, \text{ в решті випадків.} \end{cases} \quad (1.5)$$

Після того, як u_t і l_t знайдені, потрібно визначати характеристики S і d :

$$S = \sum S_t, \quad (1.6)$$

$$d = \sum d_t, \quad (1.7)$$

де

$$S_t = u_t + l_t, \quad (1.8)$$

$$d_t = u_t + l_t. \quad (1.9)$$

Підсумовування у формулах (1.6) і (1.7) проводиться по всіх членах ряду.

Неважко бачити, що S_t приймає значення 0 і 1: $S_t = 0$ у випадку, якщо x_t не є ні найбільшим, ні найменшим значенням серед всіх попередніх, інакше $S_t = 1$. Величина S знаходиться в межах $0 \le S \le n-1$. Якщо всі значення змінних рівні (нульова дисперсія), то $S = 0$, якщо ж вони монотонно ростуть або зменшуються, або коливання їх чергуються систематично збільшуючись або падаючи, то $S = n-1$.

В свою чергу величина d_t приймає значення 0 , 1 и -1. Межі для d дорівнюють: нижня складає $-(n-1)$, верхня $+(n-1)$. Нижня межа відповідає тому, що монотонно убуває, а верхня – ряду, що монотонно росте. Слабкість методу криється за ситуації, коли d рівне 0.

Що стосується першого випадку, то він відповідає повній відсутності тренда. Другий же момент може спостерігатися і тоді, коли ряд охоплює два періоди з протилежними тенденціями тренда. Крім того, $d = 0$ і у разі, коли підйоми і падіння значень $x(t)$ чергуються. Якщо змінна симетрично розташовується біля горизонтальної лінії, то величина $d = 0$ дійсно відповідає відсутності тренда в середній. Проте при визначенні d не беруться до уваги величини відхилень від горизонтальної лінії. Тому імовірна така ситуація, при якій відхилення з одним знаком будуть систематично вище за відхилення з іншим знаком. В цьому випадку зростання середнього значення (падіння) не відіб'ються на величині d. Ця ситуація на практиці, мабуть, зустрічається украй рідко, але треба мати на увазі те, що все-таки вона можлива.

Показники S і d асимптотичні нормальні і мають незалежні розподіли. Вони істотно залежать від порядку розташування змінних в часі. Показник S застосовується для виявлення тенденцій зміни дисперсії, d - для виявлення тренда в середній. Після того, як для досліджуваного ряду знайдені фактичні значення d і, S перевіряється гіпотеза про те, чи можна рахувати випадкові різниці $(d-0)$ і $(S-M_x)$ Гіпотези можна перевірити, застосовуючи t - критерій Стьюдента, тобто

$$\hat{t} = \frac{S-\mu}{\hat{\sigma}_t}, \tag{1.10}$$

$$\hat{t} = \frac{d-0}{\hat{\sigma}_2}, \tag{1.11}$$

де μ - математичне очікування величини, визначуване для випадкового розташування змінних в часі;

$\hat{\sigma}_1$ - середня квадратична помилка величини S;

$\hat{\sigma}_2$ - середня квадратична помилка величини d.

По Ф. Фостеру і А. Стюарту середні квадратичні помилки рівні:

$$\sigma_1 = \sqrt{2\sum_{t=2}^{n}\frac{1}{t} - 4\sum_{t=2}^{n}\frac{1}{t^2}} \approx \sqrt{2\ln n - 3.1253}; \tag{1.12}$$

$$\sigma_2 = \sqrt{2\sum_{t=2}^{n}\frac{1}{t}} \approx \sqrt{2\ln n - 0.8456}. \tag{1.13}$$

Значення μ, σ_1 і σ_2 табульовані, табл.1.1.

Таблиця 1.1

n	μ	σ_1	σ_2

10	3,858	1,288	1,964
15	4,636	1,521	2,153
20	5,195	1,677	2,279
25	5,632	1,791	2,373
30	5,990	1,882	2,447
35	6,264	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645
55	7,187	2,163	2,681
60	7,360	2,201	2,713
65	7,519	2,236	2,742
70	7,666	2,268	2,769
75	7,803	2,297	2,793
80	7,931	2,324	2,816
85	8,051	2,349	2,837
90	8,165	2,373	2,857
95	8,273	2,395	2,876
100	8,375	2,416	2,894

Приклад:
Дано: 1. Ряд спостережень за солоністю (табл.1.2) у термін 1954-1969 р., ‰.
Потрібно: 1. Виявити наявність тренда середнього та дисперсії.

Таблиця 1.2

Рік	x_t	u_t	l_t	Рік	x_t	u_t	l_t
1954	10,3	0	0	1962	17,1	0	0
55	14,3	1	0	63	7,7	0	0
56	7,7	0	1	64	15,3	0	0
57	15,8	1	0	65	16,3	0	0
58	14,4	0	0	66	19,9	0	0
59	16,7	1	0	67	14,4	0	0
60	15,3	0	0	68	18,7	0	0
61	20,2	1	0	69	20,7	1	0

На основі приведених даних набудемо $S=5+1=6, d=5-1=4$. Значення μ , σ_2 і σ_2 знайдемо по табл. 1.1 інтерполяцією. Отримаємо μ =4.749; σ_1 =1.552; σ_2=2.178. Звідси при перевірці d :

$$\hat{t}=\frac{4}{2.178}=1.84; \qquad (1.14)$$

при перевірці S

$$\hat{t} = \frac{(6-4.749)}{1.552} = 0.81; \qquad (1.15)$$

При рівні значущості $\alpha = 0.10$ табличне значення $t_{\frac{\alpha}{2}} \approx 1.7$.

Таким чином, гіпотеза про відсутність тренда середнього відхиляється. Що стосується гіпотези про відсутність тренда дисперсії, то вона не відхиляється, оскільки $\hat{t}_S < t_{\frac{\alpha}{2}}$.

Метод, заснований на, d - критерії і S - критерії вельми простий і легко може застосовуватися в практичних розробках. Проте слід мати на увазі, що методи різниці середніх значень і Фостера-Стюарта достатньо точні і строгі тільки в тому випадку, якщо розподіли величин нормальні.

1.1.3 Критерії серій

Вищеописані методи розроблені в припущенні, що початкові випадкові величини, на яких добуваються вибірки, підкоряються нормальному закону. На практиці ці методи з достатнім ступенем точності можна застосовувати, ймовірно, і для аналізу багатьох випадкових величин, що не підкоряються нормальному розподілу. Проте не існує достатньо чіткого критерію, користуючись яким можна було визначити, наскільки може відрізнятися розподіл випадкової величини від гаусовського (нормального), щоб ці методи були не застосовні. Ці утруднення можна обійти, якщо використовувати статистичні методи, в яких не вводяться ніяких припущень про конкретний вид розподілу початкової випадкової величини, що вивчається. Такі методи називаються не залежними від форми розподілу, або ***непараметричними.***

Критерій серій не залежить від форми розподілу випадкових величин; дозволяє встановити залежність (незалежність) послідовних значень ряду один від одного, тобто встановити присутність тренда.

Розглянемо послідовність N спостережених значень випадкової величини x і кожне спостережене значення віднесемо до однієї з двох категорій, що взаємно виключаються, які можна позначити просто знаками плюс (+) і мінус (-). Наприклад, послідовність зміряних значень, $i = 1,2,3...N$, середнє значення яких рівне $\bar{x}$. Кожне спостережене значення $x_i \geq \bar{x}$ (+) або $x_i < \bar{x}$ (-). Послідовність знаків може виглядати таким чином:

$$\underbrace{+\ +}_{1}\ \underset{2}{-}\ \underbrace{+\ +}_{3}\ \underset{4}{-}\ \underbrace{+\ +\ +}_{5}\ \underset{6}{-}\ \underset{7}{+}\ \underbrace{-\ -}_{8}\ \underset{9}{-}\ \underbrace{-\ -}_{10}\ \underset{11}{+}\ \underbrace{-\ -\ -}_{12}$$

Серією називають послідовність спостережених значень, перед якими і після яких розташовані спостережене значення іншої категорії або спостереження відсутні взагалі. У даному прикладі є r = 12 серій в послідовності N = 20 спостережень.

Число серій, які зустрічаються в послідовності спостережень, дозволяє визначити, чи є результати незалежними випадковими над однією і тією ж випадковою величиною. Зокрема, якщо послідовність N спостережень є незалежними спостереженими значеннями однієї і тієї ж випадкової величини, тобто вірогідність знаків (+) і (-) не міняється від одного спостереження до іншого, то вибірковий розподіл серій в послідовності є випадкова величина r з середнім значенням

$$\mu_2 = \frac{2N_1N_2}{N} + 1 \qquad (1.16)$$

і дисперсією

$$\sigma_r^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)} \qquad (1.17)$$

Тут - число спостережень із знаком (+), - число спостережень із знаком (-). У окремому випадку, коли $= = N_1 / N_2$ N 2, співвідношення (1.12) і (1.13) приймають вигляд

$$\mu_2 = \frac{N}{2} + 1 \qquad (1.18)$$

$$\sigma_r^2 = \frac{N(N-2)}{4(N-1)} \qquad (1.19)$$

Хай є підстави вважати, що послідовність спостережених значень містить тренд. Це означає, що вірогідність знаків (+) і (-) міняється від одного спостереженого значення до іншого. Наявність тренда можна перевірити таким чином. Введемо в розгляд гіпотезу про відсутність тренда, тобто припустимо, що послідовність спостережених значень містить тільки незалежні значення однієї і тієї ж випадкової величини. Вважаючи тепер, що число спостережених значень із знаком (+) рівне числу значень із знаком (-), можна вважати, що число серій в послідовності матиме деяку функцію розподілу (табл. 5 додаток). Гіпотезу про відсутність тренда можна тепер перевірити при будь-якому заданому рівні значущості шляхом зіставлення спостереженого числа серій з граничними значеннями $r_{n;1-\frac{\alpha}{2}}$ і $r_{n;\frac{\alpha}{2}}$, де $n = \frac{N}{2}$. Якщо спостережене число серій виходить за межі цього інтервалу, гіпотезу слід відкинути при рівні значущості α. У противному разі її слід прийняти.

Приклад:

Припустимо, що послідовність N =20 спостережених значень температури води t °С складається з наступних чисел:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
t°C	5,5	5,1	5,7	5,2	4,8	5,7	5,0	6,5	5,4	5,8	6,8	6,6	4,9	5,4	5,9	5,4	6,8	5,8	6,9	5,5

Визначимо, чи є незалежними спостережені значення $t\,^\circ$С шляхом перевірки чисел серій, які зустрічаються, якщо відлічувати спостережені значення від їх медіани (рівною 5,6°С). Виконаємо перевірку при рівні значущості α = 0,05.

Вважатимемо, що спостережені значення $t\,^\circ$С зверху 5,6°С (мають знак (+), а менш 5,6°С (знак (-). В результаті маємо послідовність:

- -	+	- -	+	-	+	-	+ + +	- -	+	-	+ + +	-
1	2	3	4	5	6	7	8	9	10	11	12	13

Таким чином, з послідовності 20 спостережених значень температури виділено 13 серій. Розглянемо гіпотезу про незалежність спостережених значень. Область ухвалення цієї гіпотези визначається інтервалом $\left[r_{10;1-\frac{\alpha}{2}} < r \le r_{10;\frac{\alpha}{2}}\right]$.

З табл. 5 (додаток), при $\alpha =$ 0,05 знаходимо $r_{10;1-\frac{\alpha}{2}} = r_{10;0.975} = 6$ і $r_{10;1-\frac{\alpha}{2}} = r_{10;0.025} = 15$. Гіпотеза приймається, оскільки r = 13 входить в інтервал між 6 і 15. Отже, немає причин ставити під сумнів ту обставину, що спостережені значення незалежні. Це означає, що немає ніяких доказів присутності тренда.

Зміст завдання

Дано: 1. Ряд спостережень елементу гідрофізичного середовища (індивідуальні дані надає викладач)

Потрібно: 1. Виконати перевірку тренду середнього та дисперсії.

1.1 Виконати перевірку способом різниці середніх

1.2 Методом Фостера – Стюарта.

1.3 Використовуючи критерії серій.

2. Скласти звіт по схемі:

а) Основні поняття.

б) Порядок виконання завдання (методика).

в) Аналіз результатів.

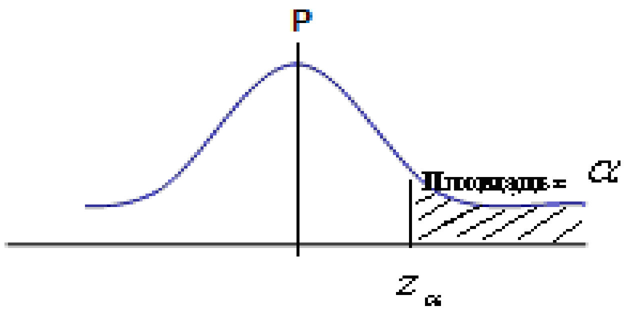
г) Виводи

д) Література

Таблиця 1

ПЛОЩІ, ЩО ПОКРИВАЮТЬСЯ ОРДИНАТАМИ НОРМОВАНОЇ ГАУСОВСЬКОЇ ЩІЛЬНОСТІ РОЗПОДІЛУ

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{\infty} e^{-\frac{z^2}{2}} dz - P[z > z_\alpha]$$



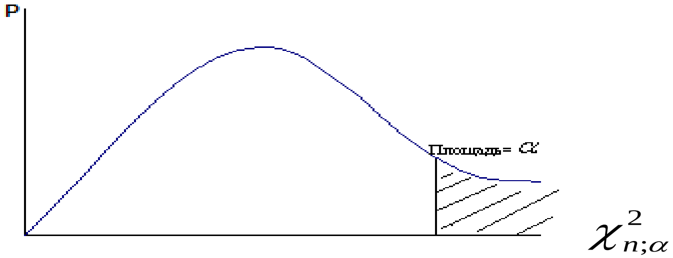
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,100,	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0466	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,00990	0,00964	0,00939	0,00914	0,00889	0,0866	0,00842
2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,0657	0,00639
2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,0289	0,00280	0,00272	0,00264
2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
2,9	0,00187	0,00181	0,00175	0,001169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139

Таблиця 2

ВІДСОТКОВІ ТОЧКИ РОЗПОДІЛУ χ^2

$$P\left[\chi_n^2 > \chi_{n;\alpha}^2\right] = \alpha$$



χ_n^2	α									
	0,995	**0,990**	**0,975**	**0,950**	**0,900**	**0,10**	**0,05**	**0,025**	**0,010**	**0,005**
1	0,000039	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,506	0,103	0,211	4,61	5,99	7,38	9,21	10,60
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,34	12,84
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,14	13,28	14,86
5	0,412	0,554	0,831	1,15	1, 61	9,24	11,07	12,83	15,09	16,75
6	0,676	0,872	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,989	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,96
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	2,21	25,19
11	2,60	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80

Продовження таблиці 2

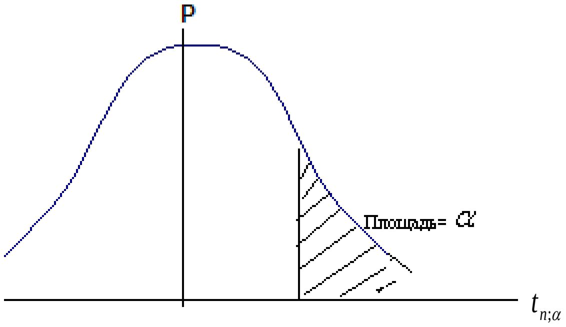
χ_n^2	α									
	0,995	**0,990**	**0,975**	**0,950**	**0,900**	**0,10**	**0,05**	**0,025**	**0,010**	**0,005**
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,2	14,61	16,47	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	49,64
28	12,46	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	53,67
40	20,71	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,59	66,77
60	35,53	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38	91,95
120	83,85	86,92	91,58	95,70	100,62	140,23	146,57	152,21	158,95	163,65

При $n > 120$ $\chi^2_{n;\alpha} \approx n\left[1 - \frac{2}{9n} + z_\alpha \sqrt{\frac{2}{9n}}\right]^3$, де z_α - задана процентна точка нормованого гаусовського розподілу .

Таблиця 3

ПРОЦЕНТНІ ТОЧКИ t - РОЗПОДІЛУ СТЬЮДЕНТА

$$P\left[t_n > t_{n;\alpha}\right] = \alpha$$



$t_{n;\alpha}$	α				
	0,10	0,050	0,025	0,010	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	20,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,767

Продовження таблиці 3

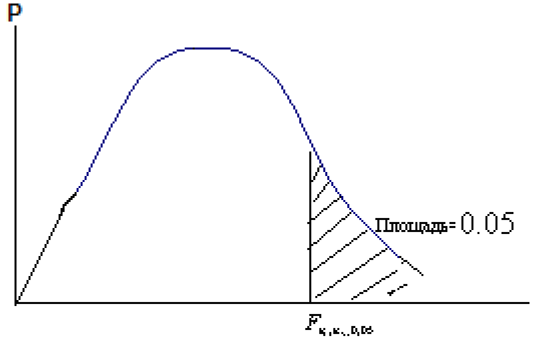
$t_{n;\alpha}$	α				
	0,10	0,050	0,025	0,010	0,005
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617

Значення α =0,995; 0,990; 0,975; 0,950; і 0,900 отримують користуючись співвідношенням $t_{n;1-\alpha} = t_{n;\alpha}$.

Таблиця 4

ПРОЦЕНТНІ ТОЧКИ F - РОЗПОДІЛУ

$$P\left[F_{n_1,n_2} > F_{n_1,n_2;0.05}\right] = 0.05$$



n2 / n1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**16**
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18,5	19,0	19,2	19,2	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,69
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,84
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,73	4,70	4,68	4,66	4,64	4,60
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,92
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,49
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,20
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	2,99
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,83

Продовження таблиці 4

n2 / n1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**16**
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,70
12	4,75	3,89	3,49	3,25	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,60
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,51
14	4,60	3,74	3,35	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,44
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,33
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,25
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,13
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,15	2,13	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,05
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,02
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,227	2,21	2,16	2,13	2,09	2,06	2,04	1,99
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,90
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,85
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,82
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,77
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,75
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,69
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,66
	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72	1,69	1,64

Продовження таблиці 4

18	**20**	**22**	**24**	**26**	**28**	**30**	**40**	**50**	**60**	**80**	**100**	**200**	**500**		n2 / n1
247	248	249	249	249	250	250	251	252	252	252	253	254	254	254	1
19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	2
8,67	8,66	8,65	8,64	8,63	8,62	8,62	8,59	8,59	8,57	8,56	8,55	8,54	8,53	8,53	3
5,82	5,80	5,79	5,77	5,76	5,75	5,75	5,72	5,70	5,69	5,67	5,66	5,65	5,64	5,63	4
4,58	3,56	4,54	4,53	4,52	4,50	44,50	4,46	4,44	4,43	4,41	4,41	4,39	4,37	4,37	5
3,90	3,87	3,86	3,84	3,83	3,82	3,81	3,77	3,75	3,74	3,72	3,71	3,69	3,68	3,67	6
3,47	3,44	3,43	3,41	3,40	3,39	3,38	3,34	3,32	3,30	3,29	3,27	3,25	3,24	3,23	7
3,17	3,15	3,13	3,12	3,10	3,09	3,08	3,04	3,02	3,01	2,99	2,97	2,95	2,94	2,93	8
2,96	2,94	2,92	2,90	2,89	2,87	2,86	2,83	2,80	2,79	2,77	2,76	2,73	2,72	2,71	9
2,80	2,77	2,75	2,74	2,72	2,71	2,70	2,66	2,64	2,62	2,60	2,59	2,56	2,55	2,54	10
2,67	2,65	2,63	2,61	2,59	2,56	2,57	2,53	2,51	2,49	2,47	2,46	2,43	2,42	2,40	11
2,57	2,54	2,52	2,51	2,49	2,48	2,47	2,43	2,40	2,38	2,36	2,35	2,32	2,31	2,30	12
2,48	2,46	2,44	2,42	2,41	2,39	2,38	2,34	2,31	2,30	2,27	2,26	2,23	2,22	2,21	13
2,41	2,38	2,37	2,35	2,33	2,32	2,31	2,27	2,24	2,22	2,20	2,19	2,16	2,14	2,13	14
2,30	2,28	2,25	2,24	2,22	2,21	2,19	2,15	2,12	2,11	2,08	2,07	2,04	2,02	2,01	16
2,22	2,19	2,17	2,15	2,13	2,12	2,11	2,06	2,04	2,02	1,99	1,98	1,95	1,93	1,92	18
2,15	2,12	2,10	2,08	2,07	2,05	2,04	1,99	1,97	1,95	1,92	1,91	1,88	1,86	1,84	20
2,10	2,07	2,05	2,03	2,01	2,00	1,98	1,94	1,91	1,89	1,86	1,85	1,82	1,80	1,78	22
2,05	2,03	2,00	1,98	1,97	1,95	1,94	1,89	1,86	1,84	1,82	1,80	1,77	1,75	1,73	24
2,02	1,99	1,97	1,95	1,93	1,91	1,90	1,84	1,82	1,80	1,78	1,76	1,73	1,71	1,69	26
1,99	1,96	1,93	1,91	1,90	1,88	1,87	1,82	1,79	1,77	1,74	1,73	1,69	1,67	1,65	28
1,96	1,93	1,91	1,89	1,87	1,85	1,84	1,79	1,76	1,74	1,71	1,70	1,66	1,64	1,62	30

Продовження таблиці 4

18	**20**	**22**	**24**	**26**	**28**	**30**	**40**	**50**	**60**	**80**	**100**	**200**	**500**		n2 / n1
1,87	1,84	1,81	1,79	1,77	1,76	1,74	1,69	1,66	1,64	1,61	1,59	1,55	1,53	1,51	40
1,81	1,78	1,76	1,74	1,72	1,70	1,69	1,63	1,60	1,58	1,54	1,52	1,48	1,46	1,44	50
1,78	1,75	1,72	1,70	1,68	1,66	1,65	1,59	1,56	1,53	1,50	1,48	1,44	1,41	1,39	60
1,73	1,70	1,68	1,65	1,63	1,62	1,60	1,54	1,51	1,48	1,45	1,43	1,38	1,35	1,32	80
1,71	1,68	1,65	1,63	1,61	1,59	1,57	1,52	1,48	1,45	1,41	1,39	1,34	1,31	1,28	100
1,66	1,62	1,60	1,57	1,55	1,53	1,52	1,46	1,41	1,39	1,35	1,32	1,26	1,22	1,19	200
1,62	1,59	1,56	1,54	1,52	1,50	1,48	1,42	1,38	1,34	1,30	1,28	1,21	1,16	1,11	500
1,60	1,57	1,54	1,52	1,501,48	1,46	1,46	1,39	1,35	1,32	1,27	1,24	1,17	1,11	1,00	

Таблиця 5

ПРОЦЕНТНІ ТОЧКИ РОЗПОДІЛУ СЕРІЙ

$$P[r_n > r_{n;\alpha}] = \alpha, \text{ де } n = N_1 = N_2 = N/2$$

$N/2$	α					
	0,99	0,975	0,95	0,05	0,025	0,01
5	2	2	3	8	9	9
6	2	3	3	10	10	11
7	3	3	4	11	12	12
8	4	4	5	12	13	13
9	4	5	6	13	14	15
10	5	6	6	15	15	16
11	6	7	7	16	16	17
12	7	7	8	17	18	18
13	7	8	9	18	19	20
14	8	9	10	19	20	21
15	9	10	11	20	21	22
16	10	11	11	22	22	23
18	11	12	13	24	25	26
20	13	14	15	26	27	28
25	17	18	19	32	33	34
30	21	22	24	37	39	40
35	25	27	28	43	44	46
40	30	31	33	48	50	51
45	34	36	37	54	55	57
50	38	40	42	59	61	63
55	43	45	46	65	66	68
60	47	49	51	70	72	74
65	52	54	56	75	77	79
70	56	58	60	81	83	85
75	61	63	65	86	88	90
80	65	68	70	91	93	96
85	70	72	74	97	99	101
90	74	77	79	102	104	107
95	79	82	84	107	109	112
100	84	86	88	113	115	117

Література

1. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов.- М.: Изд. Мир, 1974.

Завдання 2

ВЗАЄМОЗВ'ЯЗОК ЧАСТОТИ І ВІРОГІДНОСТІ ПОДІЙ

Основні поняття.

Вірогідність події є чисельна міра ступеня її можливості. Подія, яка в ході досліду, відбувається обов'язково – називається достовірною. Протилежне їй – подія неможлива; протилежна подія в ході цього ж досліду ніколи не відбудеться. При цьому під подією в теорії вірогідності розуміється всякий факт, який в результаті досліду може відбутися або не відбутися. Достовірній події приписують вірогідність рівну одиниці, неможлива подія має нульову вірогідність. Інші події – можливі, але не достовірні – характеризуються вірогідністю меншою одиниці, яка складає певну частку одиниці.

Таким чином, діапазон зміни вірогідності будь-яких подій – числа від 0 до 1. Якщо оцінюється можливість появи події у відсотках, то достовірна подія володіє 100%-й вірогідністю, неможлива подія має нульову вірогідність. Безпосередній підрахунок вірогідності можливий тільки для повної групи несумісних і рівно можливих подій.

Декілька подій в конкретному досвіді утворюють повну групу подій, якщо в результаті досвіду неодмінно повинно з'явитися хоч би одне з них. Події називаються несумісними, якщо ніякі два з них не можуть з'явитися разом.

Декілька подій в даному досліді називаються рівно можливими, якщо за умовами симетрії є підстави вважати, що не одна з цих подій не є об'єктивно більш можливою, ніж інші. Події, створюючи таку повну групу, називаються в математиці і статистиці "випадками" (у теорії вірогідності "шансами"). Про такий дослід говорять, що він "зводиться до схеми випадків". Безпосередній підрахунок вірогідності в даному випадку заснований на оцінці частки так званих "сприятливих" випадків в загальному числі випадків.

Випадок називається сприятливим, або що "сприяє" деякій події, якщо поява цього випадку спричиняє за собою появу даної події.

Якщо дослід зводиться до схеми випадків, то вірогідність події А в даному досліді можна оцінити по відносній частці сприятливих випадків. Вірогідність події A ***обчислюється як відношення числа сприятливих випадків до загального числа випадків:***

$$P(A) = \frac{m}{n}, \qquad (1)$$

де $P(A)$- вірогідність події A; n- загальне число випадків, які складають повну групу подій;
m - число випадків, сприятливих події A.

Класична формула (1) використовується для обчислення вірогідності тоді (і лише тоді), коли ***дослід зводиться до схеми випадків, тобто володіє***

симетрією можливих результатів несумісних подій утворюючих повну групу.

В даний час при ***визначенні вірогідності*** зазвичай виходять з інших принципів, ***безпосередньо пов'язуючи поняття вірогідності з емпіричним поняттям частоти події.*** Розглянемо, наприклад, неправильно виконану, несиметричну гральну кість. Випадання визначень грані вже не характеризуватиметься вірогідністю 1/6; разом з тим ясно, що для даної конкретної несиметричної кості випадання цієї грані має деяку вірогідність, яка вказує наскільки часто в середньому повинна з'являтися дана грань при багатократному киданні. Очевидно, що вірогідність таких подій не може бути обчислена за точною математичною формулою (1), оскільки відповідні цьому випадку досліди не зводяться до схеми випадків. Разом з тим ясно, що кожна з перерахованих подій володіє певним ступенем об'єктивної можливості, яку в принципі можна зміряти чисельно і яка при повторенні подібних дослідів відбиватиметься у відносній частоті відповідних подій. Тому ми вважатимемо, що кожна подія, пов'язана з масою однорідних дослідів, що зводяться до схеми випадків, має певну вірогідність, ув’язнену між нулем і одиницею. Для подій, що зводяться до схеми випадків, ця вірогідність може бути обчислена безпосередньо за формулою (1). ***Для подій, що не зводяться до схеми випадків, застосовуються інші способи визначення вірогідності.*** Всі ці способи корінням своїми йдуть у філософію, в досвід, в експеримент, в спостереження природних явищ, і для того, щоб скласти уявлення про ці способи, необхідно чітко з'ясувати собі поняття частоти події і специфіку того органічного зв'язку, тій аналогії, яка існує між вірогідністю і частотою.

Якщо проводиться серія з n дослідів, в кожному з яких могло з'явитися або не з'явитися деяка подія A, то частотою події A в даній серії дослідів називається відношення числа дослідів, в яких з'явилася подія A, до загального числа проведених дослідів.

Частоту події часто називають його “статистичною вірогідністю”, на відміну від раніше введеної “математичної вірогідності”, (1).

Позначатимемо частоту (статистичну вірогідність) події A знаком $P^*(A)$. Частота події обчислюється на підставі результату досліду (спостереження***)*** по формулі

$$P^*(A) = \frac{m^*}{n^*}, \tag{2}$$

де m^* число появи події A в досліді;

n^* - загальне число проведених дослідів, яке складає тільки деяку частину “гіпотетичної (здогадної) повної групи подій”.

При невеликому числі дослідів (спостережень випадкового явища) частота події m^* носить значною мірою випадковий характер і може помітно змінюватися від однієї групи до іншої. Наприклад, при якихось десяти киданнях монети цілком можливо, що герб з'явиться тільки двічі (частота появи герба буде рівна 0,2); при інших десяти киданнях ми цілком можемо отримати 8 гербів (частота 0,8). Проте, при збільшенні числа дослідів частота події все більш втрачає свій випадковий характер; випадкові обставини, властиві кожному окремому досвіду, в масі своїй взаємно погашаються, і частота проявляє тенденцію стабілізуватися, наближаючись з незначними коливаннями до деякої середньої, постійної величини. Наприклад, при достатньо багатократному киданні монети частота появи герба буде лише трохи ухилятися від $\frac{1}{2}$.

Ця властивість ***"стійкості частот"*** одна з найбільш характерних закономірностей, спостережуваних у випадкових явищах. Математичне формулювання цієї закономірності вперше дав Я. Бернуллі в своїй теоремі, яка є простою формою закону великих чисел. Я. Бернуллі довів, що при необмеженому збільшенні числа однорідних незалежних дослідів з практичною достовірністю можна стверджувати, що частота події буде скільки завгодно мало відрізнятися від його вірогідності в окремому досліді (п.13.5,[1]).

Зв'язок між частотою події і її вірогідністю – глибокий, природний зв'язок. Ці два поняття по суті не розділити. Дійсно, коли ми оцінюємо ступінь можливості якої-небудь події, ми неминуче пов'язуємо цю оцінку з більшою або меншою частотою появи аналогічних подій на практиці. Характеризуючи вірогідність події яким-небудь числом, ми не можемо додати цьому числу іншого реального значення і іншого практичного сенсу, чим відносна частота появи даної події при великому числі дослідів. Чисельна оцінка ступеня можливості події за допомогою вірогідності має практичний сенс саме тому, що вірогідніші події відбуваються в середньому частіше, ніж менш вірогідні.

І якщо оброблювані спостереження гідрометеорологічних елементів безумовно указують на те, що при збільшенні числа досліджених даних частота події має тенденцію вирівнюватися, наближаючись крізь ряд випадкових ухилень до деякого постійного чисельного значення, природно припустити, що це число і є вірогідність події.

Перевірити таке припущення ми можемо тільки для таких подій, вірогідність яких ми можемо обчислити безпосередньо по (1), тобто для подій, , що зводяться до схеми випадків, оскільки тільки для цих подій існує точний спосіб обчислення математичної вірогідності. Численні досліди, що проводилися протягом 300 років, дійсно підтверджують це припущення. Вони показують, що для події, що зводиться до схеми випадків, частота події при збільшенні числа дослідів завжди наближається до його вірогідності. Цілком природно припустити, що і для природних випадкових явищ (подій), що не зводяться до схеми випадків, той же закон залишається в силі і що постійне значення, до якого при збільшенні числа спостережень наближається частота події, є не що інше, як вірогідність події. Тоді частоту події при достатньо великому числі спостережень можна прийняти за наближене значення вірогідності. Так і поступають при обробці гідрометеорологічних спостережень, визначаючи з спостережених даних вірогідність подій, що не зводяться до схеми випадків. Залишається, правда, відкритим питання про "достатнє число даних спостережень", не обмежуючись утвердженням – "обробці піддалися всі дані, які накопичені до даного моменту".

Слід зазначити, що в математичній статистиці характер наближення частоти до вірогідності при збільшенні числа дослідів декілька відрізняється від "прагнення до межі" в математичному сенсі слова.

Коли в математиці затверджується, що змінна X_n із зростанням n прагне до постійної межі, то це означає, що різниця $|X_n - a|$ стає менше будь-якого додатного числа ε для всіх значень n, починаючи з деякого достатньо великого числа.

Щодо частоти події і його вірогідності такого категоричного ствердження зробити не можна. Дійсно, немає нічого фізично неможливого в тому, що при великому числі дослідів частота події значно ухилятиметься від його вірогідності; але таке значне ухилення є вельми мало вірогідним, тим менш вірогідним, ніж більше число дослідів проведене. Наприклад, при киданні монети 10 разів фізично можливо, що все 10 разів з'явиться герб, і частота появи герба буде рівна 1; при 1000 киданнях така подія все ще залишається фізично можливою, але набуває настільки малої вірогідності, що його можна рахувати, практично не здійсненним. Таким чином, при зростанні числа дослідів частота наближається до вірогідності, але не з повною достовірністю, а з більшою вірогідністю, яка при достатньо великому числі дослідів може розглядатися як практична достовірність.

У теорії вірогідності надзвичайно часто такий характер наближення одних величин до інших, і для його опису введений спеціальний термін: "збіжність по вірогідності".

Свідчать, що величина X_n сходиться по вірогідності до величини a, якщо при скільки завгодно малому ε вірогідність нерівності $|X_n - a| < \varepsilon$ із збільшенням n необмежено наближається до одиниці.

Застосовуючи цей термін, можна сказати, що при збільшенні числа дослідів (об'єму вибірки) ***частота події (статистична вірогідність) не "прагне" до вірогідності події (математичній вірогідності) а "сходиться до неї по вірогідності".***

Ця властивість частоти і вірогідності, викладена тут на підставі здорового глузду, складає зміст теореми Бернулі: при необмеженому збільшенні числа дослідів n частота події $P^*(A)$ сходиться по вірогідності до її вірогідності $P(A)$. Доведення теореми приведене в (п.13.5[1]).

Таким чином, користуючись зв'язком (аналогією) між частотою і вірогідністю, дослідник дістає можливість приписати визначення вірогідності, ув'язнені між нулем і одиницею, не тільки подіям, які зводяться до схеми випадків, але і тим подіям, які до цієї схеми не зводяться. У останньому випадку вірогідність події може бути приблизно визначена по частоті події при великому числі дослідів (спостережень).

При статистичному аналізі океанологічної інформації ми аналізуємо кінцеві реалізації. Замість "подій" аналізуються "випадкові величини". Часто замість достовірних і неможливих подій використовуються поняття "Практично неможливі" і "практично достовірні" події. При цьому практично неможливою подією називається подія, вірогідність якої не в точності рівна нулю, але вельми близька до нуля. А практично достовірною подією називається така, вірогідність якої не в точності рівна одиниці, але вельми близька до одиниці. При цьому використовується так званий "принцип практичної упевненості", який можна сформулювати таким чином:

Якщо вірогідність деякої події A в даному досліді вельми мала, то можна бути практично упевненим в тому, що при одноразовому виконанні досліду подія A не відбудеться.

У звичайних життєвих ситуаціях ми безперервно користуємося принципом практичної упевненості, Наприклад, виїжджаючи в інститут на автобусі, ми всю свою поведінку організовуємо, не зважаючи на можливість жорстокої аварії, хоча деяка, вельми мала, вірогідність такої події завжди є.

Принцип практичної упевненості не може бути доведений математичними засобами: він підтверджується всім практичним дослідом людства.

Питання про те, наскільки мала повинна бути вірогідність події, щоб його можна було рахувати практично неможливим, виходить за рамки математичної теорії і у кожному окремому випадку вирішується з практичних міркувань відповідно до тієї важливості, яку має для нас бажаний результат досліду.

Випадковою величиною називається величина, яка в результаті досліду може прийняти те або інше значення, невідомо заздалегідь – яке саме. Випадкові величини, які можна перерахувати, називаються переривчастими або дискретними випадковими величинами. У інших випадках можливі значення випадкових величин безперервно заповнюють деякий проміжок. Такі випадкові величини називаються безперервними випадковими величинами.

"Класична" теорія вірогідності оперувала переважно з подіями. Сучасна теорія вірогідності вважає за краще, де тільки можливо, оперувати з випадковими величинами. Перехід від випадкових величин до подій (і навпаки) здійснюється за деякою попередньою угодою.

Наприклад, при побудові гістограм за даними спостережень за яким-небудь океанологічним елементом призначаються непересічні межі для визначення частоти (статистичній вірогідності) попадання спостережених величин в ці градації. Призначені межі це і є "попередня угода". Якщо спостережена випадкова величина потрапила в конкретну градацію – маємо факт достовірної події. Вірогідність достовірної події рівна 1, не потрапила –

маємо випадок неможливої події, її вірогідність рівна 0. Визначаючи на підставі всього наявного об'єму вибірки число попадань в заданий інтервал, ми можемо скористатися формулою (2) і визначити вірогідність попадання океанологічної характеристики, що вивчається, в заданий інтервал. Це буде статистична вірогідність (частота), яка при зміні об'єму вибірки може мати інше числове значення, оскільки відомо, що частота прагне до вірогідності по вірогідності. При великому об'ємі вибірки вибіркова частота стає майже не випадковою величиною, практично достовірною.

Крім того, при практичному застосуванні імовірнісних методів дослідження океанологічних процесів завжди необхідно усвідомлювати, чи дійсно досліджуване явище належить до категорії масових явищ, для яких, принаймні, на деякій ділянці часу, виконується властивість стійкості частот. Тільки в цьому випадку має сенс говорити про вірогідність подій, зважаючи на не математичні фікції, а реальні характеристики випадкових явищ.

Статистичні закономірності спостерігаються завжди, коли ми маємо справу з масою однорідних випадкових явищ. Закономірності, що виявляються в цій масі, виявляються практично незалежними від індивідуальних особливостей окремих випадкових явищ, що входять в масу. Ці окремі фізичні особливості в масі як би взаємно гасяться, нівелюються, і в середньому результат маси випадкових явищ виявляється практично вже не випадковим. Виникає питання, коли "маса" стає достатньою для отримання не випадкових результатів?

Ясно, що не випадковий результат можливий тільки при нескінченно великому об'ємі однорідної вибірки. Реальні вибіркові дані (ряди спостережень) завжди кінцеві (обмежені в часі). Тому слід користуватися правилом, що відносна помилка обчислення частоти (статистичній вірогідності) не повинна перевищувати відносної точності початкових даних (точність вимірювання гідрофізичних характеристик). При цьому слід пам'ятати про одну з початкових вимог теорії вірогідності: різні випадкові величини володіють однаковою можливістю появи. Це властивість випадкових величин є одним з найделікатніших пунктів теорії вірогідності [2].

Зміст завдання

Дано:1.Ряд спостережень елементу гідрофізичного середовища (індивідуальні дані надає викладач) .

Потрібно:

1). Скласти ряд за рангом.

2). Призначити межі непересічних 10 інтервалів при однаковому числі випадків попадання в кожен інтервал.

3). Розчленувати початковий ряд на вибірки різних об'ємів: 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, >100.

4). Визначити частоту попадання в кожен інтервал при різних об'ємах початкового ряду.

5). Побудувати графіки залежності статистичної вірогідності від довжини вибірки.

6). Визначити особливості збіжності частоти до вірогідності від довжини вибірки.

7). Скласти звіт по схемі:

а) Основні поняття.
б) Порядок виконання завдання (методика).
в) Аналіз результатів.
г) Виводи.
д) Література.

Література

1. Вентцель Е.С. Теория вероятностей. “Наука”, 1969.
2. Лаплас П.С. Опыт философии теории вероятностей,1908.

Завдання 3

МЕТОД НАЙБІЛЬШОЇ ПРАВДОПОДІБНОСТІ (МНП)

1. Основні поняття й ідеї методу.

Завдання одержання достовірних статистичних оцінок океанологічних процесів є однієї з головних. Ці оцінки характеризують основні імовірнісні властивості процесів: математичне очікування, асиметрію, дисперсію й інші їхні особливості. В інших випадках - числові значення параметрів законів розподілу. Оцінки визначаються на основі вибірок, як правило, обмеженого обсягу, тому містять у собі деякі помилки. Помилки будуть найменшими, якщо оцінки задовольняють властивостям несумісності, ефективності й ґрунтовності. Одним зі способів одержання оцінок з такими властивостями, за даними вибірок, є метод найбільшої правдоподібності (МНП), називаний іноді методом максимальної правдоподібності. Цей метод розроблений видатним англійським математиком Р.А. Фішером. Ідея методу полягає в наступному:

Нехай $\boldsymbol{x}_1 , \boldsymbol{x}_2 ,\ldots \boldsymbol{x}_n$ – ряд спостережень над випадковою величиною $\boldsymbol{x}$ обсягом $\boldsymbol{n}$;

$f(x,\theta)$ - закон розподілу цієї випадкової величини залежить, як бачимо, від параметра θ. Параметр θ потрібно визначити по вибірці. Найбільш придатне обране значення θ (оцінка параметра θ) відповідає максимуму функції найбільшої правдоподібності, що має вигляд:

$$L = f(x_1,\theta)\cdot f(x_2,\theta)\cdot\ldots f(x_i,\theta)\cdot\ldots f(x_n,\theta) \tag{1.1}$$

Якщо функція L диференціюєма, то для знаходження оцінки найбільшої правдоподібності необхідно вирішити рівняння:

$$\partial L/\partial\theta \qquad (1,2)$$

і відібрати те рішення, що перетворює функцію L в максимум.

Одне з достоїнств використання функції правдоподібності при аналізі даних спостережень складається так само в тім, що часто удається знайти графічний вид цієї функції й тим самим, наочно представити результати: зобразивши графічно функцію $L(\theta)$ можна відразу знайти $\theta = \theta$, якому відповідає «найвища» крапка кривої. У тому випадку, коли є два невідомих параметри θ і φ, функція правдоподібності являє собою поверхню, морфологічні особливості якої можна зобразити відповідними ізолініями на площині (θ,φ) і по них легко можна знайти

$$L = f(\theta,\varphi/x_1)\cdot f(\theta,\varphi/x_2)\cdot\ldots f(\theta,\varphi/x_i)\cdot\ldots f(\theta,\varphi/x_n) \tag{1.3}$$

крапку максимальної правдоподібності. Звичайно, коли максимум розташований усередині припустимої області, контурами рівної правдоподібності є замкнуті криві. Помітимо, що в цьому випадку двомірним довірчим інтервалом служить площа, охоплена контуром – ізолінією значення L. Цей метод незручний, якщо є три або більша кількість параметрів, оцінки яких залежні. Метод максимальної правдоподібності не завжди дозволяє знайти оцінки, що задовольняють всім вищезгаданим вимогам. Наприклад, застосування МНП до Гаусовському розподілу, приводить до оцінки математичного очікування у вигляді

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.4}$$

середнього арифметичного ряду, яке, як відомо, є ґрунтовною, незміщеною, ефективною оцінкою свого параметра (математичного очікування). На основі МПН і при нормальному розподілі вихідного ряду, дисперсія генеральної сукупності повинна оцінюватися співвідношенням:

$$D_a = \frac{1}{n}\sum_{i=1}^{n}\left(a_i - \bar{x}\right)^2 . \tag{1.5}$$

Формула (1.5) дає ґрунтовну оцінку дисперсії, але не завжди є ефективною, внаслідок заміни математичного очікування його оцінкою (середнім арифметичним) та виявляється зміщеною на величину $\frac{n}{n-1}$, яку називають поправкою Беселя

1.2 Зміст завдання

Дано: 1) Ряд спостережень гідрофізичної характеристики, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,…$\boldsymbol{x}_i$,… $\boldsymbol{x}_n$ (індивідуальні дані надає викладач)

2) Методичні вказівки

Потрібно: 1) Скласти функцію максимальної правдоподібності Гаусовського розподілу

2) Розробити алгоритм і скласти програму розрахунку значень функції $\boldsymbol{L}\left(m_x, \sigma_x^2\right)$; по заданих m_x, σ_x^2 і відомих значеннях x_i ;

3) Обчислити значення $\boldsymbol{L}\left(m_x, \sigma_x^2\right)$ й побудувати графічне зображення $\boldsymbol{L}$ в поле (m_x, σ_x^2);

4) Провести ізолінії значень L і уточнити розрахунком шукані m_x й σ_x^2 ;

5) Скласти звіт по схемі:

а) Основні поняття.

б) Порядок виконання завдання (методика).

в) Аналіз результатів.
г) Виводи
д) Література

Методичні рекомендації до виконання п. 2:

Записати функцію найбільшої правдоподібності у вигляді:

$$L=\left\{\frac{1}{\sigma_x\sqrt{2\pi}}e^{\frac{\left(x_1-m_x\right)^2}{2\sigma_x^2}}\right\}\left\{\frac{1}{\sigma_x\sqrt{2\pi}}e^{\frac{(x_2-m_x)^2}{2\sigma_x^2}}\right\}\text{ДО}\left\{\frac{1}{\sigma_x\sqrt{2\pi}}e^{\frac{(x_n-m_x)^2}{2\sigma_x^2}}\right\}=$$

$$\left\{\frac{1}{\sigma_x\sqrt{2\pi}}\right\}^n e^{-\frac{1}{2\sigma_x^2}\left\{(x_1-m_x)^2-\kappa(x_n-m_x)^2\right\}}=\left\{\frac{1}{\sigma_x\sqrt{2\pi}}\right\}^n e^{-\frac{1}{2\sigma_x^2}\sum(x_i-m_x)^2}. \quad (1.7)$$

Для спрощення пошуку екстремумів має сенс шукати максимум не самої функції, а $ln\,L$, оскільки й L і її логарифм мають максимум в одній і тій же точці,

$$Ln\,L = n\,ln\left(\frac{1}{\sigma_x\sqrt{2\pi}}\right)-\frac{1}{2\sigma_x^2}\sum(x_i-m_x)^2 . \quad (1.8)$$

Для знаходження оцінок параметрів m_x і σ_x^2 необхідно вирішити спільно наступну систему рівнянь, вважаючи, що L диференціюємо відносно m_x й σ_x^2

$$\begin{cases}\dfrac{\partial \ln L}{\partial\ m_x}=0\\[2ex] \dfrac{\partial \ln L}{\partial\ \left(\sigma_x^2\right)}=0\end{cases}. \quad (1.9)$$

Аналізуючи результати, отримані при рішенні системи (1.9), потрібно враховувати, що другі похідні й визначник системи повинні задовольняти відомим умовам:

$$\frac{\partial^2 \ln L}{\partial\ m_x^2<0};\qquad \frac{\partial^2 \ln L}{\partial\ \left(\sigma_x^2\right)^2}<0;\qquad \text{Det}=\begin{vmatrix}\dfrac{\partial^2 \ln L}{\partial\ m_x^2}; & \dfrac{\partial^2 \ln L}{\partial\ m_x^2\left(\sigma_x^2\right)^2};\\[2ex] \dfrac{\partial^2 \ln L}{\partial\ \left(\sigma_x^2\right)^2\partial m_x} & \dfrac{\partial^2 \ln L}{\partial\ \left(\sigma_x^2\right)^2};\end{vmatrix}>0\ . \quad (1.10)$$

При виконанні самостійної роботи варто орієнтуватися на одержання проміжних результатів:

$$\frac{\partial \ln L}{\partial m_X} = \frac{\sum_{i=1}^{n}(x_i - m_X)}{\sigma_X^2}; \frac{\partial \ln L}{\partial (\sigma_X^2)} = -\frac{n}{2\sigma_X^2} + \frac{\sum_{i=1}^{n}(x_1 - m_X)^2}{2\sigma_X^4};$$

$$\frac{\partial 2\ln L}{\partial m_X \sigma (\sigma_X^2)} = \frac{\partial^2 \ln L}{\partial (\sigma_X^2)\sigma m_X} = -\frac{\sum_{i=1}^{n}(x_i - m_X)}{\sigma_X^4}. \tag{1.11}$$

Література:

1. Ван дер Варден (Русский перевод: Математическая статистика, М., ИЛ, 1960.).
2. Худсон Д. «Статистика для физиков», Изд. «Мир», 1967.